GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 2.55496 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886
Perfect score: 429
Sequence: 1 LAGQTHEQQHTTLLALVRSH.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	253	59.0	3739	9	US-09-860-846-33	Sequence 33, Appl
2	253	59.0	3739	9	US-09-988-384B-33	Sequence 33, Appl
3	253	59.0	3739	9	US-09-836-821-33	Sequence 33, Appl
4	253	59.0	3739	9	US-09-793-708-2	Sequence 2, Appli
5	253	59.0	3739	10	US-09-861-289-33	Sequence 33, Appl
6	253	59.0	11877	9	US-09-860-846-6	Sequence 6, Appli
7	253	59.0	11877	9	US-09-836-821-6	Sequence 6, Appli
8	253	59.0	11877	10	US-09-861-289-6	Sequence 6, Appli
9	253	59.0	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	240	55.9	3816	9	US-09-808-880-3	Sequence 3, Appli
11	236	55.0	4150	9	US-09-808-880-2	Sequence 2, Appli
12	234	54.5	1562	9	US-09-860-846-35	Sequence 35, Appl
13	234	54.5	1562	9	US-09-988-384B-35	Sequence 35, Appl
14	234	54.5	1562	9	US-09-836-821-35	Sequence 35, Appl
15	234	54.5	1562	9	US-09-793-708-3	Sequence 3, Appli
16	234	54.5	1562	10	US-09-861-289-35	Sequence 35, Appl
17	229	53.4	3519	9	US-09-808-880-4	Sequence 4, Appli
18	224	52.2	5215	9	US-09-860-846-2	Sequence 2, Appli
19	224	52.2	5215	9	US-09-988-384B-2	Sequence 2, Appli
20	224	52.2	5215	9	US-09-836-821-2	Sequence 2, Appli
21	224	52.2	5215	10	US-09-861-289-2	Sequence 2, Appli
22	209	48.7	4551	9	US-09-793-708-1	Sequence 1, Appli
23	209	48.7	4613	9	US-09-860-846-31	Sequence 31, Appl
24	209	48.7	4613	9	US-09-988-384B-31	Sequence 31, Appl
25	209	48.7	4613	9	US-09-836-821-31	Sequence 31, Appl
26	209	48.7	4613	10	US-09-861-289-31	Sequence 31, Appl
27	199	46.4	1616	9	US-09-712-363-262	Sequence 262, App
28	191	44.5	1346	9	US-09-793-708-4	Sequence 4, Appli
29	189	44.1	1346	9	US-09-860-846-37	Sequence 37, Appl
30	189	44.1	1346	9	US-09-988-384B-37	Sequence 37, Appl
31	189	44.1	1346	9	US-09-836-821-37	Sequence 37, Appl
32	189	44.1	1346	10	US-09-861-289-37	Sequence 37, Appl
33	178	41.5	1402	9	US-09-712-363-166	Sequence 166, App
34	164	38.2	3798	9	US-10-014-717-6	Sequence 6, Appli
35	164	38.2	7257	9	US-10-014-717-5	Sequence 5, Appli
36	111	25.9	352	10	US-09-073-009-14	Sequence 14, Appl
37	111	25.9	352	10	US-09-023-588-14	Sequence 14, Appl
38	111	25.9	352	10	US-09-793-306-14	Sequence 14, Appl
39	105.5	24.6	1421	9	US-10-014-717-2	Sequence 2, Appli
40	101.5	23.7	1832	9	US-10-014-717-4	Sequence 4, Appli
41	93.5	21.8	2439	9	US-10-014-717-7	Sequence 7, Appli
42	90	21.0	1812	10	US-09-775-938A-38	Sequence 38, Appl
43	88	20.5	1827	9	US-09-712-363-261	Sequence 261, App
44	75.5	17.6	90	9	US-09-976-059-12	Sequence 12, Appl
45	67	15.6	517	9	US-09-738-626-4534	Sequence 4534, Ap

ALIGNMENTS

RESULT 1
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33

Query Match 59.0%; Score 253; DB 9; Length 3739;
Best Local Similarity 62.2%; Pred. No. 3.6e-21;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

```
Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641
```

RESULT 2
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             59.0%;  Score 253;  DB 9;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 3.6e-21;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 3
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             59.0%;  Score 253;  DB 9;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 3.6e-21;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 4
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-2

  Query Match             59.0%;  Score 253;  DB 9;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 3.6e-21;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 5
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             59.0%;  Score 253;  DB 10;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 3.6e-21;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
```

```
                ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match             59.0%;  Score 253;  DB 9;  Length 11877;
  Best Local Similarity   62.2%;  Pred. No. 1.6e-20;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590


RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-6

  Query Match             59.0%;  Score 253;  DB 9;  Length 11877;
  Best Local Similarity   62.2%;  Pred. No. 1.6e-20;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590


RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             59.0%;  Score 253;  DB 10;  Length 11877;
  Best Local Similarity   62.2%;  Pred. No. 1.6e-20;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590


RESULT 9
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             59.0%;  Score 253;  DB 9;  Length 12199;
  Best Local Similarity   62.2%;  Pred. No. 1.6e-20;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       8831 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8890

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
```

```
Db       8891 GLRLPATLVFDHPTPGELAGHL 8912


RESULT 10
US-09-808-880-3
; Sequence 3, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

  Query Match             55.9%;  Score 240;  DB 9;  Length 3816;
  Best Local Similarity   67.2%;  Pred. No. 1.4e-19;
  Matches   45;  Conservative    9;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         16 LVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNP 75
              ||| |:| ||||: |: |  ::||||:||||||||||||||:  |||||| || ||:|||
Db       3640 LVRGHVALVLGHSGPEAIDAEQAFRDIGFDSLTAVELRNRLNAETGLRLPGTLVFDYPNP 3699

Qy         76 TTLTHHL 82
              : |  ||
Db       3700 SALADHL 3706


RESULT 11
US-09-808-880-2
; Sequence 2, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-2

  Query Match             55.0%;  Score 236;  DB 9;  Length 4150;
  Best Local Similarity   57.3%;  Pred. No. 4.6e-19;
  Matches   47;  Conservative    9;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| |  :|   |: |||   | ||||:  : : |||||:||||||||:|||||||:  |
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVAPDRAFKDLGFDSLTSVELRNRLNTAT 2500

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| |  ||:  |  |  ||
Db       2501 GLRLPVTAVFDYARPAALAGHL 2522


RESULT 12
US-09-860-846-35
; Sequence 35, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-35

  Query Match             54.5%;  Score 234;  DB 9;  Length 1562;
  Best Local Similarity   58.7%;  Pred. No. 2.3e-19;
  Matches   44;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
              ::   || |||:| | ||||::|| : | ||| :||||||||:|||:||   | ||| |
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LAFDHPNPTTLTHHL 82
              |||| |  |  ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 13
US-09-988-384B-35
; Sequence 35, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
```

```
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-35

  Query Match             54.5%;  Score 234;  DB 9;  Length 1562;
  Best Local Similarity   58.7%;  Pred. No. 2.3e-19;
  Matches   44;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
              ::   || |||:| | ||||::|| : | ||| :||||||||||:|||:||   | ||| |
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LAFDHPNPTTLTHHL 82
                ||||  |  |  ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 14
US-09-836-821-35
; Sequence 35, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-35

  Query Match             54.5%;  Score 234;  DB 9;  Length 1562;
  Best Local Similarity   58.7%;  Pred. No. 2.3e-19;
  Matches   44;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
              ::   || |||:| | ||||::|| : | ||| :||||||||||:|||:||   | ||| |
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LAFDHPNPTTLTHHL 82
                ||||  |  |  ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 15
US-09-793-708-3
; Sequence 3, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-3

  Query Match             54.5%;  Score 234;  DB 9;  Length 1562;
  Best Local Similarity   58.7%;  Pred. No. 2.3e-19;
  Matches   44;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
              ::   || |||:| | ||||::|| : | ||| :||||||||||:|||:||   | ||| |
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LAFDHPNPTTLTHHL 82
                ||||  |  |  ||
Db       1460 TVFDHPTPAALAAHL 1474


Search completed: June 17, 2003, 13:46:48
Job time : 3.55496 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 1.49039 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886
Perfect score: 429
Sequence: 1 LAGQTHEQQHTTLLALVRSH.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	270	62.9	3729	2	US-08-804-227C-4	Sequence 4, Appli
2	256	59.7	3724	2	US-08-804-227C-10	Sequence 10, Appl
3	256	59.7	3724	2	US-08-804-198-4	Sequence 4, Appli
4	253	59.0	3739	3	US-09-320-878-2	Sequence 2, Appli
5	253	59.0	3739	4	US-09-105-537-33	Sequence 33, Appl
6	253	59.0	11877	4	US-09-105-537-6	Sequence 6, Appli
7	240	55.9	3816	4	US-09-428-517-3	Sequence 3, Appli
8	236	55.0	4150	4	US-09-428-517-2	Sequence 2, Appli
9	235	54.8	3567	2	US-07-642-734C-4	Sequence 4, Appli
10	235	54.8	3567	3	US-08-439-009A-4	Sequence 4, Appli
11	234	54.5	1562	3	US-09-320-878-3	Sequence 3, Appli
12	234	54.5	1562	4	US-09-105-537-35	Sequence 35, Appl
13	229	53.4	3519	4	US-09-428-517-4	Sequence 4, Appli
14	226	52.7	3170	4	US-09-036-987A-4	Sequence 4, Appli
15	226	52.7	3170	4	US-09-370-700-4	Sequence 4, Appli
16	224	52.2	5215	4	US-09-105-537-2	Sequence 2, Appli
17	223	52.0	1996	2	US-08-804-227C-9	Sequence 9, Appli
18	223	52.0	1996	2	US-08-804-198-3	Sequence 3, Appli
19	223	52.0	3491	2	US-07-642-734C-2	Sequence 2, Appli
20	223	52.0	3491	3	US-08-439-009A-2	Sequence 2, Appli
21	219	51.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
22	217	50.6	4545	2	US-08-804-227C-14	Sequence 14, Appl
23	217	50.6	4550	2	US-08-804-227C-8	Sequence 8, Appli
24	217	50.6	4550	2	US-08-804-198-2	Sequence 2, Appli
25	216	50.3	4928	4	US-09-036-987A-5	Sequence 5, Appli
26	216	50.3	4928	4	US-09-370-700-5	Sequence 5, Appli
27	214	49.9	502	4	US-09-413-814-106	Sequence 106, App
28	213.5	49.8	1864	2	US-08-804-227C-3	Sequence 3, Appli
29	209	48.7	4551	3	US-09-320-878-1	Sequence 1, Appli
30	209	48.7	4613	4	US-09-105-537-31	Sequence 31, Appl
31	208	48.5	5588	4	US-09-036-987A-6	Sequence 6, Appli
32	208	48.5	5588	4	US-09-370-700-6	Sequence 6, Appli
33	207	48.3	2152	4	US-09-036-987A-3	Sequence 3, Appli
34	207	48.3	2152	4	US-09-370-700-3	Sequence 3, Appli
35	207	48.3	3170	2	US-07-642-734C-5	Sequence 5, Appli
36	207	48.3	3170	3	US-08-439-009A-5	Sequence 5, Appli
37	206.5	48.1	185	4	US-09-154-083-14	Sequence 14, Appl
38	203	47.3	1611	2	US-08-804-227C-5	Sequence 5, Appli
39	198	46.2	1891	2	US-08-804-227C-12	Sequence 12, Appl
40	198	46.2	1891	2	US-08-804-198-6	Sequence 6, Appli
41	191	44.5	1346	3	US-09-320-878-4	Sequence 4, Appli
42	189	44.1	1346	4	US-09-105-537-37	Sequence 37, Appl
43	185	43.1	2595	4	US-09-036-987A-2	Sequence 2, Appli
44	185	43.1	2595	4	US-09-370-700-2	Sequence 2, Appli
45	180	42.0	6095	4	US-09-144-085-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-4
; Sequence 4, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3729 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-4

Query Match 62.9%; Score 270; DB 2; Length 3729;
Best Local Similarity 63.4%; Pred. No. 5.6e-25;
Matches 52; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

```
Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              | | | :: |  :|| :|:| | |||| : |:|| ||||:||||||||||||:|||||  |
Db       3547 LTGLTAQEGHALVLAEIRAHAAAVLGHGSDDSIPEDRAFKDLGFDSLTAVEMRNRLSAAT 3606

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
```

```
                 ||||| || |||| |  |  ||
Db       3607 GLRLPATLVFDHPTPGELAGHL 3628


RESULT 2
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-10

  Query Match             59.7%;  Score 256;  DB 2;  Length 3724;
  Best Local Similarity   58.5%;  Pred. No. 3.3e-23;
  Matches   48;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| : |::|  || ||  |:| |||| :   : ||| ||::||||||||||||::  |
Db       3548 LAGLSAEERHDRLLGLVGEHVAAVLGHGSAAEVRPDRPFREVGFDSLTAVELRNRMAAVT 3607

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| || |||| |  |: ||
Db       3608 GVRLPATLVFDHPTPAALSSHL 3629


RESULT 3
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-4

  Query Match             59.7%;  Score 256;  DB 2;  Length 3724;
  Best Local Similarity   58.5%;  Pred. No. 3.3e-23;
  Matches   48;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| : |::|  || ||  |:| |||| :   : ||| ||::||||||||||||::  |
Db       3548 LAGLSAEERHDRLLGLVGEHVAAVLGHGSAAEVRPDRPFREVGFDSLTAVELRNRMAAVT 3607

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| || |||| |  |: ||
Db       3608 GVRLPATLVFDHPTPAALSSHL 3629


RESULT 4
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-2
```

```
  Query Match             59.0%;  Score 253;  DB 3;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 8e-23;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 5
US-09-105-537-33
; Sequence 33, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-33

  Query Match             59.0%;  Score 253;  DB 4;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 8e-23;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 6
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             59.0%;  Score 253;  DB 4;  Length 11877;
  Best Local Similarity   62.2%;  Pred. No. 3.8e-22;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590


RESULT 7
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3

  Query Match             55.9%;  Score 240;  DB 4;  Length 3816;
  Best Local Similarity   67.2%;  Pred. No. 3.7e-21;
  Matches   45;  Conservative    9;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         16 LVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNP 75
              ||| |:| ||||: |: |  ::||||:|||||||||||||:  |||||| || ||:|||
Db       3640 LVRGHVALVLGHSGPEAIDAEQAFRDIGFDSLTAVELRNRLNAETGLRLPGTLVFDYPNP 3699

Qy         76 TTLTHHL 82
              : |  ||
Db       3700 SALADHL 3706


RESULT 8
US-09-428-517-2
; Sequence 2, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
```

```
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-2

  Query Match             55.0%;  Score 236;  DB 4;  Length 4150;
  Best Local Similarity   57.3%;  Pred. No. 1.3e-20;
  Matches   47;  Conservative    9;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| |  :|   |: |||   | ||||:  : : |||||:||||||||:|||||||:  |
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVAPDRAFKDLGFDSLTSVELRNRLNTAT 2500

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |||||  |  ||:  |  |  ||
Db       2501 GLRLPVTAVFDYARPAALAGHL 2522


RESULT 9
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3567 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-642-734C-4

  Query Match             54.8%;  Score 235;  DB 2;  Length 3567;
  Best Local Similarity   56.1%;  Pred. No. 1.5e-20;
  Matches   46;  Conservative   10;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||::   |   |  ||||| | | |: : | :|  :||:||||||| ||||||||   |
Db       3403 LAGRSESDQVAGLAELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTAT 3462

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||:|| |||| |  :  ||
Db       3463 GVRLPSTLVFDHPTPLAVAEHL 3484


RESULT 10
US-08-439-009A-4
; Sequence 4, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3567 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-439-009A-4

  Query Match             54.8%;  Score 235;  DB 3;  Length 3567;
  Best Local Similarity   56.1%;  Pred. No. 1.5e-20;
  Matches   46;  Conservative   10;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||::   |   |  ||||| | | |: : | :|  :||:||||||| ||||||||   |
Db       3403 LAGRSESDQVAGLAELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTAT 3462

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||:|| |||| |  :  ||
Db       3463 GVRLPSTLVFDHPTPLAVAEHL 3484


RESULT 11
US-09-320-878-3
; Sequence 3, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
```

```
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-3

  Query Match             54.5%;  Score 234;  DB 3;  Length 1562;
  Best Local Similarity   58.7%;  Pred. No. 6.5e-21;
  Matches   44;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
              ::   || |||:| | ||||::|| : | ||| :|||||||||:|||:||   | ||| |
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LAFDHPNPTTLTHHL 82
                |||| |  |  ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 12
US-09-105-537-35
; Sequence 35, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-35

  Query Match             54.5%;  Score 234;  DB 4;  Length 1562;
  Best Local Similarity   58.7%;  Pred. No. 6.5e-21;
  Matches   44;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
              ::   || |||:| | ||||::|| : | ||| :|||||||||:|||:||   | ||| |
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LAFDHPNPTTLTHHL 82
                |||| |  |  ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 13
US-09-428-517-4
; Sequence 4, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 3519
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-4

  Query Match             53.4%;  Score 229;  DB 4;  Length 3519;
  Best Local Similarity   53.7%;  Pred. No. 8.3e-20;
  Matches   44;  Conservative   14;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |:| :  :|   |: |||:  | |||| :   :| :|||::|||||||||||||| |:  |
Db       1479 LSGLSRVRQEEELVELVRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAAT 1538

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| |:|||||| | :   |
Db       1539 GIRLPATMAFDHPNATAIARFL 1560


RESULT 14
US-09-036-987A-4
; Sequence 4, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  3170 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-4

  Query Match             52.7%;  Score 226;  DB 4;  Length 3170;
  Best Local Similarity   56.1%;  Pred. No. 1.7e-19;
  Matches   46;  Conservative   11;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||      |:   |: || : :|||||| : : : |:|||| |||||| ||:|||||:  |
Db       1458 LAELPEAQRRHELVDLVCAQVATVLGHGSREEVQPERAFRALGFDSLMAVDLRNRLTTAT 1517

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |||||||  ||:|||  |  ||
Db       1518 GLRLPTTTVFDYPNPAALAAHL 1539


RESULT 15
US-09-370-700-4
; Sequence 4, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 3170
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-4

  Query Match             52.7%;  Score 226;  DB 4;  Length 3170;
  Best Local Similarity   56.1%;  Pred. No. 1.7e-19;
  Matches   46;  Conservative   11;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||      |:   |: || : :|||||| : : : |:|||| |||||| ||:|||||:  |
Db       1458 LAELPEAQRRHELVDLVCAQVATVLGHGSREEVQPERAFRALGFDSLMAVDLRNRLTTAT 1517

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |||||||  ||:|||  |  ||
Db       1518 GLRLPTTTVFDYPNPAALAAHL 1539


Search completed: June 17, 2003, 13:09:31
Job time : 1.49039 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 4.01696 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886
Perfect score: 429
Sequence: 1 LAGQTHEQQHTTLLALVRSH.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	429	100.0	3972	21	AAB23749	S. avermitilis ave
2	429	100.0	3972	22	AAG65264	Streptomyces averm
3	429	100.0	3972	22	AAG65268	Streptomyces averm
4	429	100.0	5532	21	AAB23752	S. avermitilis ave
5	429	100.0	5532	22	AAG65267	Streptomyces averm
6	270	62.9	3729	18	AAW22603	Tylactone synthase
7	256	59.7	3724	18	AAW23718	Platenolide syntha
8	256	59.7	3724	18	AAW22608	Platenolide syntha
9	254	59.2	4881	21	AAB23751	S. avermitilis ave
10	254	59.2	4881	22	AAG65266	Streptomyces averm
11	253	59.0	3739	21	AAB18638	Amino acid sequenc
12	253	59.0	3739	21	AAY77193	S. venezuelae macr
13	253	59.0	3739	21	AAY77201	S. venezuelae pik
14	253	59.0	3739	21	AAY67202	Narbonolide syntha
15	253	59.0	12199	21	AAY77180	S. venezuelae pik
16	246	57.3	1721	19	AAW52847	A. mediterranei ri
17	245	57.1	6239	21	AAB23750	S. avermitilis ave
18	245	57.1	6239	22	AAG65265	Streptomyces averm
19	241	56.2	6797	22	AAB31558	Pimaricin biosynth
20	240	55.9	3816	21	AAY92708	S. antibioticus 8,
21	239	55.7	1561	21	AAY77202	S. venezuelae pik
22	237	55.2	1688	19	AAW52848	A. mediterranei ri
23	236	55.0	899	22	AAE10153	Protein encoded by
24	236	55.0	4150	21	AAY92707	S. antibioticus 8,
25	236	55.0	11096	22	AAE10129	Streptomyces nours
26	235	54.8	3567	14	AAR44431	eryA region polype
27	235	54.8	9477	22	AAE10144	Streptomyces nours
28	234	54.5	1560	21	AAY67203	Narbonolide syntha
29	234	54.5	1562	21	AAB18639	Amino acid sequenc
30	234	54.5	1562	21	AAY77194	S. venezuelae macr
31	233	54.3	3562	22	AAB82213	Polyketide synthas
32	233	54.3	5069	19	AAW52846	A. mediterranei ri
33	229	53.4	3519	21	AAY92709	S. antibioticus 8,
34	229	53.4	5435	22	AAE10145	Streptomyces nours
35	226	52.7	3170	20	AAY39299	SpnC a polyketide
36	226	52.7	3170	22	AAB70967	S. spinosa protein
37	224	52.2	1891	19	AAW52844	Amycolatopsis medi
38	224	52.2	3201	22	AAB82214	Polyketide synthas
39	224	52.2	3413	19	AAW52849	A. mediterranei ri
40	224	52.2	4630	21	AAY77177	S. venezuelae vep
41	223	52.0	1996	18	AAW22607	Platenolide syntha
42	223	52.0	1996	18	AAW23717	Platenolide syntha
43	223	52.0	3398	14	AAR44430	eryA region polype
44	222	51.7	3192	22	AAE10128	Streptomyces nours
45	220	51.3	3546	22	AAB82212	Polyketide synthas

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 429; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 3805 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 3864

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 3865 GLRLPTTLAFDHPNPTTLTHHL 3886

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 429; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 3805 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 3864

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 3865 GLRLPTTLAFDHPNPTTLTHHL 3886

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 429; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3805 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 3864

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||||||||||||||||||||
Db       3865 GLRLPTTLAFDHPNPTTLTHHL 3886


RESULT 4
AAB23752
ID   AAB23752 standard; Protein; 5532 AA.
XX
AC   AAB23752;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 278-305; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;

  Query Match             100.0%;  Score 429;  DB 21;  Length 5532;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-46;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 1815

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||||||||||||||||||||
Db       1816 GLRLPTTLAFDHPNPTTLTHHL 1837


RESULT 5
AAG65267
ID   AAG65267 standard; Protein; 5532 AA.
XX
AC   AAG65267;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 7.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 217-235; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   5532 AA;

  Query Match             100.0%;  Score 429;  DB 22;  Length 5532;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-46;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 1815

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||||||||||||||||||||
Db       1816 GLRLPTTLAFDHPNPTTLTHHL 1837


RESULT 6
AAW22603
ID   AAW22603 standard; Protein; 3729 AA.
XX
AC   AAW22603;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase ORF3 protein.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin.
XX
```

```
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "encoded by GTG"
FT   Domain          43..465
FT                   /note= "ketosynthase domain, KS4"
FT   Domain          588..916
FT                   /note= "acyltransferase domain, AT4"
FT   Domain          1178..1356
FT                   /note= "ketoreductase domain, KR'4"
FT   Domain          1451..1534
FT                   /note= "acyl carrier protein domain, ACP4"
FT   Domain          1556..1980
FT                   /note= "ketosynthase domain, KS5"
FT   Domain          2095..2425
FT                   /note= "acyltransferase domain, AT5"
FT   Domain          2451..2636
FT                   /note= "dehydratase domain, DH2"
FT   Domain          3274..3453
FT                   /note= "ketoreductase domain, KR5"
FT   Domain          3548..3631
FT                   /note= "acyl carrier protein domain, ACP5"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80413.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 9; Pages 86-98; 220pp; English.
XX
CC   AAW22601-W22605 represent proteins encoded by the tylactone synthase
CC   gene cluster of the invention. The gene cluster is also referred to as
CC   the tylG gene, and was isolated from Streptomyces fradiae. These
CC   sequences are multifunctional proteins which direct the synthesis of the
CC   polyketide tylactone, isolated from Streptomyces fradiae. Tylactone is
CC   the basic building block of the antibiotic tylosin. The DNA sequence can
CC   be modified so as to alter the type of carboxylic acids incorporated,
CC   the number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence   3729 AA;

  Query Match             62.9%;  Score 270;  DB 18;  Length 3729;
  Best Local Similarity   63.4%;  Pred. No. 1.6e-25;
  Matches   52;  Conservative    9;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              | | | :: |  :|| :|:| | |||| : |:|| ||||:||||||||||:|||||  |
Db       3547 LTGLTAQEGHALVLAEIRAHAAAVLGHGSDDSIPEDRAFKDLGFDSLTAVEMRNRLSAAT 3606

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3607 GLRLPATLVFDHPTPGELAGHL 3628


RESULT 7
AAW23718
ID   AAW23718 standard; Protein; 3724 AA.
XX
AC   AAW23718;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase ORF3 protein.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "encoded by TGT"
FT   Domain          36..459
FT                   /note= "ketosynthase domain, KS4"
FT   Domain          594..921
FT                   /note= "acyltransferase domain, AT4"
FT   Domain          1177..1350
FT                   /note= "ketosynthase domain, KS04"
FT   Domain          1459..1523
FT                   /note= "acyl carrier protein domain, ACP4"
FT   Domain          1545..1969
FT                   /note= "ketosynthase domain, KS5"
FT   Domain          2088..2424
FT                   /note= "acyltransferase domain, AT5"
FT   Domain          2445..2619
FT                   /note= "dehydratase domain, DH5"
FT   Domain          2912..3261
FT                   /note= "enoylreductase domain, ER5"
FT   Domain          3266..3451
FT                   /note= "ketoreductase domain, KR5"
FT   Domain          3551..3631
FT                   /note= "acyl carrier protein domain, ACP5"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   N-PSDB; AAT78508.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 8; Pages 54-66; 81pp; English.
XX
CC   AAW23716-W23720 represent proteins encoded by the platenolide synthase
CC   gene cluster of the invention. The gene cluster is also referred to as
CC   the srmG gene, and was isolated from Streptomyces ambofaciens. These
CC   sequences are multi-functional proteins which direct the synthesis of
CC   the polyketide platenolide. Platenolide is the basic building block of
CC   the macrolide antibiotic spiramycin. The DNA can be used to produce
CC   compounds exhibiting antibiotic activity based on the platenolide
CC   structure, including specifically the macrolide antibiotic spiramycin and
CC   spiramycin analogues and derivatives. Modifications of the platenolide
CC   synthase DNA sequence can be made so as to change the number and type of
CC   carboxylic acids incorporated into the growing polyketide chain and to
CC   change the kind of post-condensation processing that is conducted.
XX
SQ   Sequence   3724 AA;

  Query Match             59.7%;  Score 256;  DB 18;  Length 3724;
```

```
  Best Local Similarity   58.5%;  Pred. No. 1.1e-23;
  Matches   48;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| : |::|  || ||  |:| |||| :   : ||| ||::||||||||||||::  |
Db       3548 LAGLSAEERHDRLLGLVGEHVAAVLGHGSAAEVRPDRPFREVGFDSLTAVELRNRMAAVT 3607

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| || |||| |  |: ||
Db       3608 GVRLPATLVFDHPTPAALSSHL 3629


RESULT 8
AAW22608
ID   AAW22608 standard; Protein; 3724 AA.
XX
AC   AAW22608;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase ORF3 protein.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "encoded by TGT"
FT   Domain          36..459
FT                   /note= "ketosynthase domain, KS4"
FT   Domain          594..921
FT                   /note= "acyltransferase domain, AT4"
FT   Domain          1177..1350
FT                   /note= "ketosynthase domain, KS04"
FT   Domain          1459..1523
FT                   /note= "acyl carrier protein domain, ACP4"
FT   Domain          1545..1969
FT                   /note= "ketosynthase domain, KS5"
FT   Domain          2088..2424
FT                   /note= "acyltransferase domain, AT5"
FT   Domain          2445..2619
FT                   /note= "dehydratase domain, DH5"
FT   Domain          2912..3261
FT                   /note= "enoylreductase domain, ER5"
FT   Domain          3266..3451
FT                   /note= "ketoreductase domain, KR5"
FT   Domain          3551..3631
FT                   /note= "acyl carrier protein domain, ACP5"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80414.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 155-167; 220pp; English.
XX
CC   AAW22606-W22610 represent proteins encoded by the platenolide synthase
CC   gene cluster. The gene cluster is also referred to as the srmG gene, and
CC   was isolated from Streptomyces ambofaciens. These sequences are
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA encoding this sequence was used along
CC   with the tylG gene (see AAT80413) to create a hybrid ORF1 sequence (see
CC   AAT80415). The tylG gene is the tylactone synthase gene cluster of the
CC   invention. The tylG sequence was isolated from Streptomyces fradiae, and
CC   encodes multifunctional proteins which direct the synthesis of the
CC   polyketide tylactone. Tylactone is the basic building block of the
CC   antibiotic tylosin. The hybrid sequence can be used to transform S.
CC   ambofaciens lacking the srmG ORF1 sequence, or S. fradiae lacking the
CC   tylG ORF1 sequence, so that they can produce polyketides. The DNA
CC   sequence can be modified so as to alter the type of carboxylic acids
CC   incorporated, the number of carboxylic acids incorporated and/or the
CC   post-condensation reactions performed, thereby resulting in novel
CC   tylosin-related polyketides.
XX
SQ   Sequence   3724 AA;

  Query Match             59.7%;  Score 256;  DB 18;  Length 3724;
  Best Local Similarity   58.5%;  Pred. No. 1.1e-23;
  Matches   48;  Conservative   12;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| : |::|  || ||  |:| |||| :   : ||| ||::||||||||||||::  |
Db       3548 LAGLSAEERHDRLLGLVGEHVAAVLGHGSAAEVRPDRPFREVGFDSLTAVELRNRMAAVT 3607

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| || |||| |  |: ||
Db       3608 GVRLPATLVFDHPTPAALSSHL 3629


RESULT 9
AAB23751
ID   AAB23751 standard; Protein; 4881 AA.
XX
AC   AAB23751;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:5.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 254-278; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
```

```
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   4881 AA;

  Query Match             59.2%;  Score 254;  DB 21;  Length 4881;
  Best Local Similarity   62.2%;  Pred. No. 2.9e-23;
  Matches   51;  Conservative    6;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||      |  |:| || :| | |||| | | : |:||||||||||||||||||:  |
Db       2525 LAGLDPAAQAETVLDLVLTHAAAVLGHGTADAVVPERAFRDLGFDSLTAVELRNRLNTAT 2584

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||| | || |||| |  |  |:
Db       2585 GLRFPRTLVFDHPRPVALAAHI 2606


RESULT 10
AAG65266
ID   AAG65266 standard; Protein; 4881 AA.
XX
AC   AAG65266;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 6.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 201-217; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   4881 AA;

  Query Match             59.2%;  Score 254;  DB 22;  Length 4881;
  Best Local Similarity   62.2%;  Pred. No. 2.9e-23;
  Matches   51;  Conservative    6;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||      |  |:| || :| | |||| | | : |:||||||||||||||||||:  |
Db       2525 LAGLDPAAQAETVLDLVLTHAAAVLGHGTADAVVPERAFRDLGFDSLTAVELRNRLNTAT 2584

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||| | || |||| |  |  |:
Db       2585 GLRFPRTLVFDHPRPVALAAHI 2606


RESULT 11
AAB18638
ID   AAB18638 standard; Protein; 3739 AA.
XX
AC   AAB18638;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of narbonolide synthase subunit 2 (PICAII).
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;  99US-0320878.
XX
PR   28-MAY-1998;  98US-0087080.
PR   22-SEP-1998;  98US-0100880.
PR   08-FEB-1999;  99US-0119139.
PR   20-MAY-1999;  99US-0134990.
PR   30-APR-1997;  97US-0846247.
PR   06-MAY-1998;  98US-0073538.
PR   28-AUG-1998;  98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 11-12; 117pp; English.
XX
CC   The present sequence represents a narbonolide synthase subunit 2
CC   (PICAII). The nucleotide sequence encoding it is used in the course of
CC   the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence   3739 AA;
```

```
  Query Match             59.0%;  Score 253;  DB 21;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 2.8e-23;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 12
AAY77193
ID   AAY77193 standard; Protein; 3739 AA.
XX
AC   AAY77193;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic enzyme PikAII, SEQ ID NO:33.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87298.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 403-415; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC   Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC   AAZ87295-Z87302.
XX
SQ   Sequence   3739 AA;

  Query Match             59.0%;  Score 253;  DB 21;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 2.8e-23;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 13
AAY77201
ID   AAY77201 standard; Protein; 3739 AA.
XX
AC   AAY77201;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster protein #2.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87318.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
```

CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77200-Y77203 and AAY80997 represent proteins encoded by the
CC macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
CC ATCC 15439, as given in figure 31.
XX
SQ Sequence 3739 AA;

Query Match 59.0%; Score 253; DB 21; Length 3739;
Best Local Similarity 62.2%; Pred. No. 2.8e-23;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 3620 GLRLPATLVFDHPTPGELAGHL 3641

RESULT 14
AAY67202
ID AAY67202 standard; protein; 3739 AA.
XX
AC AAY67202;
XX
DT 23-MAR-2000 (first entry)
XX
DE Narbonolide synthase subunit 2 (PICAII) protein sequence.
XX
KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 2;
KW PICAII; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS Streptomyces venezuelae.
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR N-PSDB; AAZ56001.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Example 2; Page 13-14; 98pp; English.
XX
CC This is the Streptomyces venezuelae narbonolide synthase subunit 2,
CC PICAII protein sequence. The invention relates to recombinant DNA
CC containing a coding sequence for a narbonolide polyketide synthase
CC (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC a series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC transferase enzyme is required for this conversion, and the desosamine
CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC DNA of the invention is used to express, in transformed cells,
CC narbonolide (or its derivatives) or other ketolides (particularly
CC hybrids), which may then be converted (e.g. by other enzymes
CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.
XX
SQ Sequence 3739 AA;

Query Match 59.0%; Score 253; DB 21; Length 3739;
Best Local Similarity 62.2%; Pred. No. 2.8e-23;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 3620 GLRLPATLVFDHPTPGELAGHL 3641

RESULT 15
AAY77180
ID AAY77180 standard; Protein; 12199 AA.
XX
AC AAY77180;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87285.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 19; Page 315-353; 438pp; English.

XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents a protein
CC encoded by the macrolide biosynthetic gene cluster (pik) from
CC Streptomyces venezuelae ATCC 15439.
XX
SQ Sequence 12199 AA;

Query Match 59.0%; Score 253; DB 21; Length 12199;
Best Local Similarity 62.2%; Pred. No. 1.3e-22;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

```
Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| |||||||||||||:  |
Db       8831 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8890

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       8891 GLRLPATLVFDHPTPGELAGHL 8912
```

Search completed: June 17, 2003, 13:02:35
Job time : 5.01696 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 3.97473 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_366_451
Perfect score: 424
Sequence: 1 LTALSPTQQQSLLLDLVRAH.....RLPVTLIFDHTTPAAVAARL 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	424	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	418	98.6	3626	2	Q9F7T9	Q9f7t9 streptomyce
3	214	50.5	6146	2	Q93HJ5	Q93hj5 streptomyce
4	204	48.1	5435	2	Q9L4X2	Q9l4x2 streptomyce
5	200	47.2	3970	2	Q93HI8	Q93hi8 streptomyce
6	199	46.9	4613	2	Q9ZGI5	Q9zgi5 streptomyce
7	199	46.9	4928	2	Q9ALM3	Q9alm3 saccharopol
8	197	46.5	3524	2	Q93H86	Q93h86 streptomyce
9	197	46.5	5644	2	Q93NX8	Q93nx8 streptomyce
10	196	46.2	3654	2	O30766	O30766 streptomyce
11	196	46.2	4150	2	Q9KIV4	Q9kiv4 streptomyce
12	195	46.0	1562	2	Q9ZGI3	Q9zgi3 streptomyce
13	193	45.5	3352	2	Q93H83	Q93h83 streptomyce
14	193	45.5	3413	2	O54593	O54593 amycolatops
15	193	45.5	4881	2	Q9S0R3	Q9s0r3 streptomyce
16	192	45.3	2152	2	Q9ALM5	Q9alm5 saccharopol
17	190	44.8	3729	2	O33956	O33956 streptomyce
18	190	44.8	6145	2	Q93H84	Q93h84 streptomyce
19	189	44.6	2100	2	O30481	O30481 streptomyce
20	189	44.6	3576	16	Q9EX53	Q9ex53 streptomyce
21	188	44.3	1191	2	Q9RHV0	Q9rhv0 streptomyce
22	188	44.3	6048	2	Q93H87	Q93h87 streptomyce
23	188	44.3	6315	2	Q9ADL6	Q9adl6 polyangium
24	186	43.9	3591	2	Q9KIE1	Q9kie1 streptomyce
25	186	43.9	3613	2	Q93HJ1	Q93hj1 streptomyce
26	185	43.6	3939	2	Q93HJ3	Q93hj3 streptomyce
27	184.5	43.5	11096	2	Q9L4W3	Q9l4w3 streptomyce
28	184	43.4	1762	2	O52546	O52546 amycolatops
29	184	43.4	1763	2	O52790	O52790 amycolatops
30	184	43.4	5060	2	O52545	O52545 amycolatops
31	184	43.4	5069	2	O52789	O52789 amycolatops
32	184	43.4	8817	2	Q53840	Q53840 polyangium
33	183	43.2	4685	2	Q93HJ2	Q93hj2 streptomyce
34	183	43.2	4735	2	O54666	O54666 amycolatops
35	182	42.9	3739	2	Q9ZGI4	Q9zgi4 streptomyce
36	182	42.9	7525	2	Q9KIE0	Q9kie0 streptomyce
37	181	42.7	1762	2	O30480	O30480 streptomyce
38	181	42.7	3170	2	Q9ALM4	Q9alm4 saccharopol
39	180	42.5	6260	2	Q54299	Q54299 streptomyce
40	180	42.5	9507	2	Q9EWA1	Q9ewa1 streptomyce
41	180	42.5	10223	2	Q54296	Q54296 streptomyce
42	179	42.2	1366	2	Q9L4W5	Q9l4w5 streptomyce
43	179	42.2	1412	2	Q93NW8	Q93nw8 streptomyce
44	179	42.2	1620	16	P96285	P96285 mycobacteri
45	179	42.2	6797	2	Q9X993	Q9x993 streptomyce

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 424; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 4.4e-38;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 425

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |||||||||||||||||||||||||
Db        426 SKATGLRLPVTLIFDHTTPAAVAARL 451


RESULT 2
Q9F7T9
ID   Q9F7T9      PRELIMINARY;      PRT;  3626 AA.
AC   Q9F7T9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Avermectin polyketide synthase (Fragment).
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC31271;
RA   Hong Y.-S., Lee J.J.;
RT   "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT   and polyketide synthase gene from Streptomyces avermitilis.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF275943; AAG09812.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT   NON_TER    3626   3626
SQ   SEQUENCE   3626 AA;  380557 MW;  6272F5F088C1A8D0 CRC64;

  Query Match             98.6%;  Score 418;  DB 2;  Length 3626;
  Best Local Similarity   98.8%;  Pred. No. 1.8e-37;
  Matches   85;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 406

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |||||||||||| ||||||||||||
Db        407 SKATGLRLPVTLILDHTTPAAVAARL 432


RESULT 3
Q93HJ5
ID   Q93HJ5      PRELIMINARY;      PRT;  6146 AA.
AC   Q93HJ5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69192.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR002155; Thiolase.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF00108; thiolase; 1.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
DR   PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   6146 AA;  637460 MW;  1D56F1108D458D00 CRC64;

  Query Match             50.5%;  Score 214;  DB 2;  Length 6146;
  Best Local Similarity   53.5%;  Pred. No. 2.4e-14;
  Matches   46;  Conservative    9;  Mismatches   27;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | |||| ::  |:||||| ||  ||       :| |  |  :|  |||||:  |||||||
Db       5980 LAALSPAERHRLVLDLVREHTAGVL----GHGSADDIDPDQAFKTLGFDSLTAVELRNRL 6035

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              || | :| ||:||| ||||: | |
Db       6036 RTATSLTVPATLVFDHPTPAALTAHL 6061


RESULT 4
Q9L4X2
ID   Q9L4X2      PRELIMINARY;      PRT;  5435 AA.
AC   Q9L4X2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysJ.
GN   NYSJ.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71767.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
```

```
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5435 AA;  562659 MW;  AA55465DF087A38C CRC64;

  Query Match             48.1%;  Score 204;  DB 2;  Length 5435;
  Best Local Similarity   55.2%;  Pred. No. 2.8e-13;
  Matches   48;  Conservative    6;  Mismatches   27;  Indels    6;  Gaps    2;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPS-ASFAHLGFDSVMGVELRNR 59
              | ||    |:   ||||||  | |||    |      :|| |:   |  |||||:| |||| |
Db       5265 LAALDAEQRHQALLDLVRTETAAVLGHSG-----ADAVPAERDFNRLGFDSLMAVELRTR 5319

Qy         60 LSKATGLRLPVTLIFDHTTPAAVAARL 86
              |: ||| ||| ||:||| || |||  |
Db       5320 LATATGARLPATLVFDHPTPDAVARHL 5346


RESULT 5
Q93HI8
ID   Q93HI8      PRELIMINARY;      PRT;  3970 AA.
AC   Q93HI8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69199.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3970 AA;  412657 MW;  A93E583FAFA68C47 CRC64;

  Query Match             47.2%;  Score 200;  DB 2;  Length 3970;
  Best Local Similarity   51.2%;  Pred. No. 5.3e-13;
  Matches   44;  Conservative    9;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              || | | :|   ||||||::   ||       :|       ||: |||||:  || ||||
Db       3805 LTGLPPAEQLRRLLDLVRSNAATVL----GHASAGAIAAEQSFSELGFDSLTAVEFRNRL 3860

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ||||||| ||:|:| ||||:|  |
Db       3861 GAATGLRLPATLVFEHPTPAALAGEL 3886


RESULT 6
Q9ZGI5
ID   Q9ZGI5      PRELIMINARY;      PRT;  4613 AA.
AC   Q9ZGI5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAI.
GN   PIKAI.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR   EMBL; AF079138; AAC69329.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4613 AA;  477281 MW;  A3BAF8D37CEC9383 CRC64;

  Query Match             46.9%;  Score 199;  DB 2;  Length 4613;
  Best Local Similarity   48.8%;  Pred. No. 8.2e-13;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db       2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2541 TRATGLQLPATLVFDHPTPLALVSLL 2566


RESULT 7
Q9ALM3
ID   Q9ALM3      PRELIMINARY;      PRT;  4928 AA.
AC   Q9ALM3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase extender modules 5-7.
GN   SPND.
OS   Saccharopolyspora spinosa.
```

```
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23263.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4928 AA;  517247 MW;  C22C6361888F865A CRC64;

  Query Match             46.9%;  Score 199;  DB 2;  Length 4928;
  Best Local Similarity   50.0%;  Pred. No. 8.9e-13;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | || | ::  |||::| | | : |  |    :|    |  :|| |||||:  |||||||
Db       4762 LAALPPAERGHLLLEVVLAETASTLGHD----SAEAVQPDRTFAELGFDSLTAVELRNRL 4817

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :  |||||| ||:||| || |:: :|
Db       4818 NAVTGLRLPPTLVFDHPTPLALSEQL 4843


RESULT 8
Q93H86
ID   Q93H86      PRELIMINARY;      PRT;  3524 AA.
AC   Q93H86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69304.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3524 AA;  366619 MW;  64A65759C461EBC4 CRC64;

  Query Match             46.5%;  Score 197;  DB 2;  Length 3524;
  Best Local Similarity   46.5%;  Pred. No. 1e-12;
  Matches   40;  Conservative   16;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :  |  ::::  :|||||:   |||       :|    |  :|  |||||:  ||||||:
Db       3352 IAGLPESEREEAVLDLVRSEVAAVL----GHSSAKTVQPDQAFGDLGFDSLTAVELRNRM 3407

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :::|||||| ||:||| ||| :| |:
Db       3408 NQATGLRLPATLVFDHPTPAVLAQRV 3433


RESULT 9
Q93NX8
ID   Q93NX8      PRELIMINARY;      PRT;  5644 AA.
AC   Q93NX8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphJ.
GN   AMPHJ.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73502.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW   Phosphopantetheine.
SQ   SEQUENCE   5644 AA;  585775 MW;  CF6A35E3BA0D2658 CRC64;

  Query Match             46.5%;  Score 197;  DB 2;  Length 5644;
  Best Local Similarity   52.9%;  Pred. No. 1.8e-12;
  Matches   46;  Conservative    8;  Mismatches   27;  Indels    6;  Gaps    2;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPS-ASFAHLGFDSVMGVELRNR 59
              |  |   |:   |: |||  |  ||   |     :|| |:   |: |||||:| |||| |
Db       5474 LAGLGAEQRHETLVGLVRQETAGVLGHSG-----ADAVPADRDFSRLGFDSLMAVELRTR 5528

Qy         60 LSKATGLRLPVTLIFDHTTPAAVAARL 86
              || |||:||| ||:||| ||||||  |
Db       5529 LSAATGVRLPSTLVFDHPTPAAVARHL 5555
```

```
RESULT 10
O30766
ID   O30766      PRELIMINARY;      PRT;  3654 AA.
AC   O30766;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase modules 4 and 5.
GN   NIDA3.
OS   Streptomyces caelestis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=36816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL-2821;
RX   MEDLINE=98053867; PubMed=9393718;
RA   Kakavas S.J., Katz L., Stassi D.;
RT   "Identification and characterization of the niddamycin polyketide
RT   synthase genes from Streptomyces caelestis.";
RL   J. Bacteriol. 179:7515-7522(1997).
DR   EMBL; AF016585; AAC46026.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001827; Antennapedia.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00032; ANTENNAPEDIA; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3654 AA;  379108 MW;  59825E49EDCD06FF CRC64;

  Query Match             46.2%;  Score 196;  DB 2;  Length 3654;
  Best Local Similarity   52.3%;  Pred. No. 1.4e-12;
  Matches   45;  Conservative    9;  Mismatches   28;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |:|||  ::   ||:|||    |||        |||     :|   ||||:  |||||||
Db       3471 LSALSDEERLDTLLELVRELVAAVLGHGSAASVASD----RAFREAGFDSLTAVELRNRL 3526

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| ||:||| || |:| ||
Db       3527 AAATGLRLPATLVFDHPTPTALAGRL 3552


RESULT 11
Q9KIV4
ID   Q9KIV4      PRELIMINARY;      PRT;  4150 AA.
AC   Q9KIV4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   8,8a-deoxyoleandolide synthase 1.
GN   OLEAI.
OS   Streptomyces antibioticus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20363406; PubMed=10908114;
RA   Shah S., Xue Q., Tang L., Carney J.R., Betlach M., McDaniel R.;
RT   "Cloning, Characterization, and Heterologous Expression of a
RT   Polyketide Synthase and P-450 Oxidase Involved in the Biosynthesis of
RT   the Antibiotic Oleandomycin.";
RL   J. Antibiot. 53:502-508(2000).
DR   EMBL; AF220951; AAF82408.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4150 AA;  435261 MW;  9383296C4C16647D CRC64;

  Query Match             46.2%;  Score 196;  DB 2;  Length 4150;
  Best Local Similarity   50.0%;  Pred. No. 1.6e-12;
  Matches   43;  Conservative   10;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:  :|: ||: |||      ||   | |  |     |  :|  |||||:  |||||||
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVA----PDRAFKDLGFDSLTSVELRNRL 2496

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||||| :||:  |||:|  |
Db       2497 NTATGLRLPVTAVFDYARPAALAGHL 2522


RESULT 12
Q9ZGI3
ID   Q9ZGI3      PRELIMINARY;      PRT;  1562 AA.
AC   Q9ZGI3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAIII.
GN   PIKAIII.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR   EMBL; AF079138; AAC69331.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
```

```
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   1562 AA;  163593 MW;  DB4BA0DAAA15B63F CRC64;

  Query Match             46.0%;  Score 195;  DB 2;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 6.4e-13;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 13
Q93H83
ID   Q93H83      PRELIMINARY;      PRT;  3352 AA.
AC   Q93H83;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA5.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69307.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002362; Antenna_beta.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00969; ANTENNA_COMP_BETA; UNKNOWN_1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3352 AA;  353395 MW;  66A5865BFD2027B7 CRC64;

  Query Match             45.5%;  Score 193;  DB 2;  Length 3352;
  Best Local Similarity   50.0%;  Pred. No. 2.7e-12;
  Matches   43;  Conservative   11;  Mismatches   28;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | ||:| :|:  ||: ||| :  ||       :| |  |:| |  |||||:  ::|| ||
Db       2968 LAALAPVEQEERLLEHVRAQSAVVL----GHASAEDVLPTAHFLELGFDSLTAIDLRRRL 3023

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              | |||||||  | ||| ||| :|  |
Db       3024 SAATGLRLPSGLAFDHPTPARLAKHL 3049


RESULT 14
```

```
O54593
ID   O54593      PRELIMINARY;      PRT;  3413 AA.
AC   O54593;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase.
GN   RIFE.
OS   Amycolatopsis mediterranei (Nocardia mediterranei).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=33910;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RA   Kim C.G., Yu T.W., Fryhle C., Handa S., Floss H.G.;
RL   J. Biol. Chem. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RA   August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,
RA   Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
RA   Floss H.G.;
RL   Chem. Biol. 5:0-0(0002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RA   August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
RA   Floss H.G.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LBG A3136;
RA   Schupp T., Toupet C., Engel N., Goff S.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF040570; AAC01714.1; -.
DR   EMBL; AJ223012; CAA11039.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3413 AA;  353729 MW;  FFA25AC4A3920AA4 CRC64;

  Query Match             45.5%;  Score 193;  DB 2;  Length 3413;
  Best Local Similarity   47.7%;  Pred. No. 2.7e-12;
  Matches   41;  Conservative   14;  Mismatches   27;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:| :|::||||:||     ||    | |   :|     :|   ||||:  |||||||
Db       1598 LAGLAPAEQEALLLDVVRTQVALVLGHAGPEAVRAD----TAFKDTGFDSLTSVELRNRL 1653

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :|:||:|| ||:||: || |:|  |
Db       1654 REASGLKLPATLVFDYPTPVALARYL 1679


RESULT 15
Q9S0R3
ID   Q9S0R3      PRELIMINARY;      PRT;  4881 AA.
AC   Q9S0R3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 4.
GN   AVEA4.
```

```
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84479.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4881 AA;  510303 MW;  F6568C1F01ACD37A CRC64;

  Query Match             45.5%;  Score 193;  DB 2;  Length 4881;
  Best Local Similarity   50.0%;  Pred. No. 4.1e-12;
  Matches   43;  Conservative    7;  Mismatches   32;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  | |  |   :||||  |  |||       ||    |  :|  |||||:  |||||||
Db       2525 LAGLDPAAQAETVLDLVLTHAAAVL----GHGTADAVVPERAFRDLGFDSLTAVELRNRL 2580

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                : ||||| | ||:|||  | |:|| :
Db       2581 NTATGLRFPRTLVFDHPRPVALAAHI 2606


Search completed: June 17, 2003, 13:07:52
Job time : 4.97473 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 0.937857 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_366_451
Perfect score: 424
Sequence: 1 LTALSPTQQQSLLLDLVRAH.....RLPVTLIFDHTTPAAVAARL 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	174	41.0	3567	1	ERY2_SACER	Q03132 saccharopol
2	173	40.8	3491	1	ERY1_SACER	Q03131 saccharopol
3	166	39.2	3519	1	OL56_STRAT	Q07017 streptomyce
4	138	32.5	3172	1	ERY3_SACER	Q03133 saccharopol
5	99	23.3	1538	1	PPSB_MYCTU	Q10978 mycobacteri
6	90.5	21.3	1774	1	MSAS_PENPA	P22367 penicillium
7	84.5	19.9	4427	1	PKSL_BACSU	Q05470 bacillus su
8	81	19.1	1876	1	PPSA_MYCTU	Q10977 mycobacteri
9	81	19.1	2505	1	FAS_RAT	P12785 rattus norv
10	78.5	18.5	1278	1	DHBF_BACSU	P45745 bacillus su
11	78	18.4	838	1	FAS_MOUSE	P19096 mus musculu
12	73.5	17.3	2511	1	FAS_CHICK	P12276 gallus gall
13	72.5	17.1	331	1	GALE_LACCA	O84903 lactobacill
14	72	17.0	740	1	PURL_LACCA	P35852 lactobacill
15	72	17.0	1392	1	LYS2_YEAST	P07702 saccharomyc
16	72	17.0	3587	1	SRF1_BACSU	P27206 bacillus su
17	70	16.5	2560	1	PPS2_BACSU	P39846 bacillus su
18	68.5	16.2	3587	1	SRF2_BACSU	Q04747 bacillus su
19	68	16.0	4273	1	PKSM_BACSU	P40872 bacillus su
20	66	15.6	6486	1	TYCC_BACBR	O30409 b tyrocidin
21	65.5	15.4	398	1	BMPB_XENLA	P30884 xenopus lae
22	65	15.3	362	1	SDHL_RAT	P09367 rattus norv
23	64.5	15.2	64	1	FAS_RABIT	P07855 oryctolagus
24	64	15.1	1419	1	LYS2_SCHPO	P40976 schizosacch
25	64	15.1	2110	1	MCAS_MYCBO	Q02251 mycobacteri
26	63.5	15.0	247	1	ATPI_MESVI	Q9mus9 mesostigma
27	63.5	15.0	332	1	GALE_STRTR	P21977 streptococc
28	63.5	15.0	454	1	CBPH_LOPAM	P37892 lophius ame
29	63	14.9	887	1	ACOC_CAEEL	Q23500 caenorhabdi
30	62.5	14.7	553	1	PUR9_SYNY3	P74741 s bifunctio
31	62	14.6	87	1	ACPY_STRAW	Q93hc3 streptomyce
32	62	14.6	160	1	RBFA_STRCO	Q9z527 streptomyce
33	61.5	14.5	741	1	LOXB_PHAVU	P27481 phaseolus v
34	61	14.4	178	1	IF3_UREPA	Q9pqr4 ureaplasma
35	61	14.4	636	1	GIDA_BRUME	Q8yjs5 brucella me
36	61	14.4	1082	1	A3B2_HUMAN	Q13367 homo sapien
37	61	14.4	3649	1	ACVS_NOCLA	P27743 nocardia la
38	60.5	14.3	247	1	ATPI_MAIZE	P17344 zea mays (m
39	60.5	14.3	247	1	ATPI_ORYSA	P12083 oryza sativ
40	60.5	14.3	247	1	ATPI_WHEAT	Q9xpt0 triticum ae
41	60.5	14.3	415	1	LOLC_NEIMA	P57061 neisseria m
42	60.5	14.3	415	1	LOLC_NEIMB	P57062 neisseria m
43	60.5	14.3	530	1	MURE_MYCLE	O69557 mycobacteri
44	60.5	14.3	692	1	GLND_CORGL	Q9x706 corynebacte
45	60	14.2	213	1	HIS5_CAUCR	Q9a231 caulobacter

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             41.0%;  Score 174;  DB 1;  Length 3567;
  Best Local Similarity   48.9%;  Pred. No. 1.7e-11;
  Matches   44;  Conservative    7;  Mismatches   27;  Indels   12;  Gaps    2;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLN----DDGNERTASDAGPSASFAHLGFDSVMGVEL 56
              |    | :  | :  | :|||:|  ||       |    || |         |  |||||:  |||
Db       3403 LAGRSESDQVAGLAELVRSHAAAVSGYGSADQLPERKA--------FKDLGFDSLAAVEL 3454

Qy         57 RNRLSKATGLRLPVTLIFDHTTPAAVAARL 86
              ||||  |||:||| ||:||| || |||  |
Db       3455 RNRLGTATGVRLPSTLVFDHPTPLAVAEHL 3484


RESULT 2
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
```

```
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             40.8%;  Score 173;  DB 1;  Length 3491;
  Best Local Similarity   44.2%;  Pred. No. 2.2e-11;
  Matches   38;  Conservative   13;  Mismatches   31;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | :|   :::  | :|||:|  |||     ||  :|     :|| || ||:  :||||||
Db       1876 LASLPAPEREEALFELVRSHAAAVLGHASAERVPAD----QAFAELGVDSLSALELRNRL 1931

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |||:||| | :|||      :|| |
Db       1932 GAATGVRLPTTTVFDHPDVRTLAAHL 1957


RESULT 3
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             39.2%;  Score 166;  DB 1;  Length 3519;
  Best Local Similarity   44.6%;  Pred. No. 1.4e-10;
  Matches   37;  Conservative   12;  Mismatches   30;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |: ||  :|:  |::||||    ||       :| |     :|  |||||:  ||||| |
Db       1479 LSGLSRVRQEEELVELVRAQAAVVL----GHGSAQDVPAERAFKELGFDSLTAVELRNGL 1534

Qy         61 SKATGLRLPVTLIFDHTTPAAVA 83
              : |||:||| |: |||    |:|
Db       1535 AAATGIRLPATMAFDHPNATAIA 1557


RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
```

```
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             32.5%;  Score 138;  DB 1;  Length 3172;
  Best Local Similarity   42.3%;  Pred. No. 2.2e-07;
  Matches   33;  Conservative   12;  Mismatches   25;  Indels    8;  Gaps    3;

Qy          7 TQQQSLLLDLVRAHTMAVLNDDGNERTASDA-GPSASFAHLGFDSVMGVELRNRLSKATG 65
              | |:  ||:   :|  |:|       :: || |    |  |||||:  | |||:| :|||
Db       2817 TSQE--LLEFTHSHVAAILG-----HSSPDAVGQDQPFTELGFDSLTAVGLRNQLQQATG 2869

Qy         66 LRLPVTLIFDHTTPAAVA 83
                | || ||:|:| |   :|
Db       2870 LALPATLVFEHPTVRRLA 2887


RESULT 5
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
```

```
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             23.3%;  Score 99;  DB 1;  Length 1538;
  Best Local Similarity   35.2%;  Pred. No. 0.0029;
  Matches   25;  Conservative   13;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          8 QQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLR 67
              ::: :| | | |    |:     |       ||| |  || ||:| | |:  ||:: |
Db       1420 KRRDMLFDHVGALAATVMGMPPTE----PLDPSAGFFQLGMDSLMSVTLQRALSESLGEF 1475

Qy         68 LPVTLIFDHTT 78
              || :::||: |
Db       1476 LPASVVFDYPT 1486


RESULT 6
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH +
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             21.3%;  Score 90.5;  DB 1;  Length 1774;
  Best Local Similarity   32.9%;  Pred. No. 0.033;
  Matches   26;  Conservative   13;  Mismatches   35;  Indels    5;  Gaps    1;

Qy          5 SPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKAT 64
              |  : :: | : :|     ||     : || |    |: | || |||| | || :|
Db       1692 SGPELKAYLDEKIRGCVAKVL-----QMTAEDVDSKAALADLGVDSVMTVTLRRQLQLTL 1746

Qy         65 GLRLPVTLIFDHTTPAAVA 83
                : :| || : | | : :|
Db       1747 KIAVPPTLTWSHPTVSHLA 1765


RESULT 7
```

```
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;

  Query Match             19.9%;  Score 84.5;  DB 1;  Length 4427;
  Best Local Similarity   30.2%;  Pred. No. 0.49;
  Matches   26;  Conservative   14;  Mismatches   37;  Indels    9;  Gaps    2;

Qy          5 SPTQQQSL----LLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | |: :::    | | : |    ||  | ||       |  :|  :| ||: |:|    :
Db       2479 SDTETETITAEALCDELTAGLAEVLYMDQNE-----IDPDEAFIDIGMDSITGLEWIKAI 2533

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :|  |  | || ::|: |    | |
Db       2534 NKQYGTSLNVTKVYDYPTTRDFAVYL 2559


RESULT 8
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
```

```
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             19.1%;  Score 81;  DB 1;  Length 1876;
  Best Local Similarity   34.1%;  Pred. No. 0.44;
  Matches   15;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy         43 FAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARL 86
              || || :|:| : :|    :  |: |  |::|:| |  ::|: |
Db       1789 FAELGLNSLMAMAIRREAEQFVGIELSATMLFNHPTVKSLASYL 1832


RESULT 9
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
```

```
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match             19.1%;  Score 81;  DB 1;  Length 2505;
  Best Local Similarity   31.2%;  Pred. No. 0.63;
  Matches   24;  Conservative    9;  Mismatches   30;  Indels   14;  Gaps    1;

Qy          8 QQQSLLLDLVRAHTMAVLNDDGNE-------------RTASDAGPSASFAHLGFDSVMG 53
              |  ::|   | |   || : ||               |  :     :| | || ||:||
Db       2095 QPHAVLSSFVLAEKKAVAHGDGEAQRDLVKAVAHILGIRDLAGINLDSSLADLGLDSLMG 2154

Qy         54 VELRNRLSKATGLRLPV 70
              ||:|  | :   | ||:
Db       2155 VEVRQILEREHDLVLPI 2171


RESULT 10
```

```
DHBF_BACSU
ID   DHBF_BACSU     STANDARD;      PRT;  1278 AA.
AC   P45745;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable serine activating enzyme.
GN   DHBF.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE OF 1-95 FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=96146538; PubMed=8550523;
RA   Rowland B.M., Taber H.W.;
RT   "Duplicate isochorismate synthase genes of Bacillus subtilis:
RT   regulation and involvement in the biosyntheses of menaquinone and
RT   2,3-dihydroxybenzoate.";
RL   J. Bacteriol. 178:854-861(1996).
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: 2,3-dihydroxybenzoate biosynthesis.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99120; CAB15186.1; -.
DR   EMBL; U26444; AAC44634.1; -.
DR   HSSP; P14687; 1AMU.
```

```
DR   SubtiList; BG11243; dhbF.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00668; Condensation; 2.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Ligase; Phosphopantetheine; Complete proteome.
FT   DOMAIN      966   1033       ACYL CARRIER (ACP).
FT   BINDING     996    996       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1278 AA;  142068 MW;  B44FBFF3FCB085B4 CRC64;

  Query Match             18.5%;  Score 78.5;  DB 1;  Length 1278;
  Best Local Similarity   28.8%;  Pred. No. 0.55;
  Matches   23;  Conservative   13;  Mismatches   33;  Indels   11;  Gaps    2;

Qy          7 TQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGL 66
              | |: :| ||  |  : :          :  |   ||  ||  |::   | :|: :  |
Db        964 TPQEEILCDLF-AEVLGL----------ARVGIDDSFFELGGHSLLAARLMSRIREVMGA 1012

Qy         67 RLPVTLIFDHTTPAAVAARL 86
               |  :  :||  | | :|| |
Db       1013 ELGIAKLFDEPTVAGLAAHL 1032


RESULT 11
FAS_MOUSE
ID   FAS_MOUSE      STANDARD;      PRT;   838 AA.
AC   P19096;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]
DE   (Fragment).
GN   FASN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Liver;
RX   MEDLINE=89149781; PubMed=2920037;
RA   Paulauskis J.D., Sul H.S.;
RT   "Structure of mouse fatty acid synthase mRNA. Identification of the
RT   two NADPH binding sites.";
RL   Biochem. Biophys. Res. Commun. 158:690-695(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13135; CAA31525.1; -.
DR   PIR; A32262; A32262.
DR   MGD; MGI:95485; Fasn.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; PARTIAL.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP.
FT   NON_TER       1      1
FT   DOMAIN      451    507       ACYL CARRIER (ACP).
FT   DOMAIN      535    838       THIOESTERASE.
FT   BINDING     484    484       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    635    635       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    808    808       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   838 AA;  91212 MW;  32C40EE5CA5AB841 CRC64;

  Query Match             18.4%;  Score 78;  DB 1;  Length 838;
  Best Local Similarity   29.9%;  Pred. No. 0.38;
  Matches   23;  Conservative   11;  Mismatches   29;  Indels   14;  Gaps    1;

Qy          8 QQQSLLLDLVRAHTMAVLNDDGNE-------------RTASDAGPSASFAHLGFDSVMG 53
              |  ::|   | |   || : ||:               |  :     :: | || ||:||
Db        428 QPHAVLSSFVLAEKKAVAHGDGDNQRDLVKAVAHILGIRDLAGINLDSTLADLGLDSLMG 487

Qy         54 VELRNRLSKATGLRLPV 70
              ||:|  | :   | ||:
Db        488 VEVRQILEREHDLVLPM 504


RESULT 12
FAS_CHICK
ID   FAS_CHICK      STANDARD;      PRT;  2511 AA.
AC   P12276;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=White leghorn; TISSUE=Liver;
RX   MEDLINE=95031085; PubMed=7944406;
RA   Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Amino-terminal blocking group and sequence of the animal fatty acid
RT   synthase.";
RL   Arch. Biochem. Biophys. 314:45-49(1994).
RN   [2]
RP   SEQUENCE OF 75-1775 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89282777; PubMed=2734291;
```

```
RA   Holzer K.P., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of chicken liver fatty acid
RT   synthase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN   [3]
RP   SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89139426; PubMed=2917973;
RA   Chirala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,
RA   Wakil S.J.;
RT   "A novel cDNA extension procedure. Isolation of chicken fatty acid
RT   synthase cDNA clones.";
RL   J. Biol. Chem. 264:3750-3757(1989).
RN   [4]
RP   SEQUENCE OF 1752-2512 FROM N.A.
RX   MEDLINE=88320436; PubMed=2842766;
RA   Yuan Z., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of DNA complementary to chicken
RT   liver fatty acid synthase mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN   [5]
RP   SEQUENCE OF 2202-2512 FROM N.A.
RX   MEDLINE=89088152; PubMed=3207710;
RA   Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT   "Characterization of a genomic and cDNA clone coding for the
RT   thioesterase domain and 3' noncoding region of the chicken liver
RT   fatty acid synthase gene.";
RL   Biochemistry 27:7778-7785(1988).
RN   [6]
RP   SEQUENCE OF 2121-2209.
RX   MEDLINE=89192401; PubMed=2648999;
RA   Huang W.-Y., Stoops J.K., Wakil S.J.;
RT   "Complete amino acid sequence of chicken liver acyl carrier protein
RT   derived from the fatty acid synthase.";
RL   Arch. Biochem. Biophys. 270:92-98(1989).
RN   [7]
RP   SEQUENCE OF 2209-2508.
RC   STRAIN=White leghorn;
RX   MEDLINE=89088151; PubMed=3207709;
RA   Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Complete amino acid sequence of the thioesterase domain of chicken
RT   liver fatty acid synthase.";
RL   Biochemistry 27:7773-7777(1988).
RN   [8]
RP   SEQUENCE OF 667-674 AND 1698-1709.
RX   MEDLINE=89323081; PubMed=2751995;
RA   Chang S.I., Hammes G.G.;
RT   "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
RT   fluorescence resonance energy transfer in chicken liver fatty acid
RT   synthase.";
RL   Biochemistry 28:3781-3788(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
CC       POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04485; AAB46389.1; -.
DR   EMBL; J03860; AAA48767.1; -.
DR   EMBL; J02839; AAA82106.1; ALT_SEQ.
DR   PIR; A33918; XYCHFA.
DR   PIR; A32015; A32015.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;
KW   Alternative splicing; Pyridoxal phosphate.
FT   INIT_MET      0      0       BETA-KETOACYL SYNTHASE.
FT   DOMAIN        1   ?411       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN      427    815       ENOYL REDUCTASE.
FT   DOMAIN     1638   1866       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     1867   2119       ACYL CARRIER (ACP).
FT   DOMAIN     2124   2180       THIOESTERASE.
FT   DOMAIN     2209   2511       ACETYLATION.
FT   MOD_RES       1      1       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    160    160       ACYL/MALONYL TRANSFERASES (BY
FT   ACT_SITE    579    579       SIMILARITY).
FT                                BETA-HYDROXYACYL DEHYDRATASE (BY
FT   ACT_SITE    877    877       SIMILARITY).
FT   NP_BIND     1674   1691       NADP (ER).
FT   BINDING     1707   1707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND     1888   1903       NADP (KR).
FT   BINDING     2157   2157       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    2308   2308       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    2481   2481       THIOESTERASE (BY SIMILARITY).
FT   VARSPLIC    2348   2348       T -> TQCFSFSLF (IN ISOFORM 1).
FT   CONFLICT      77     78       QL -> PV (IN REF. 2).
FT   CONFLICT     116    116       L -> A (IN REF. 2).
FT   CONFLICT     675    675       R -> S (IN REF. 2).
FT   CONFLICT    1169   1169       K -> N (IN REF. 2).
FT   CONFLICT    1178   1178       A -> T (IN REF. 2).
FT   CONFLICT    1191   1191       R -> H (IN REF. 2).
FT   CONFLICT    1198   1198       P -> L (IN REF. 2).
FT   CONFLICT    1286   1287       DN -> ND (IN REF. 2).
FT   CONFLICT    1372   1372       K -> E (IN REF. 2).
FT   CONFLICT    1533   1533       C -> Y (IN REF. 2).
FT   CONFLICT    1577   1577       W -> R (IN REF. 2).
FT   CONFLICT    1685   1696       QAAIAIALSMGC -> ASSHCHRLEHGLA
FT                                 (IN REF. 2; AAA48767).
FT   CONFLICT    1732   1732       Q -> E (IN REF. 2).
FT   CONFLICT    1745   1745       S -> N (IN REF. 2).
SQ   SEQUENCE   2511 AA;  274648 MW;  622039DAC8315D3F CRC64;

  Query Match           17.3%;  Score 73.5;  DB 1;  Length 2511;
  Best Local Similarity  27.9%;  Pred. No. 4.6;
  Matches   19;  Conservative   12;  Mismatches   24;  Indels   13;  Gaps    1;
```

```
Qy          8 QQQSLLLDLVRAHTMAVLNDDGNE------------RTASDAGPSASFAHLGFDSVMGV 54
              |   ::   | |  ::| :: |::             |  |     :| | || ||:|||
Db       2102 QPHPVMSSFVLAEKVSVKSEGGSQRDLVEAVAHILGVRDVSSLNAESSLADLGLDSLMGV 2161

Qy         55 ELRNRLSK 62
              |:|  | :
Db       2162 EVRQTLER 2169


RESULT 13
GALE_LACCA
ID   GALE_LACCA     STANDARD;      PRT;   331 AA.
AC   O84903;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-
DE   galactose 4-epimerase).
GN   GALE.
OS   Lactobacillus casei.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1582;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=64H;
RX   MEDLINE=98268934; PubMed=9603808;
RA   Bettenbrock K., Alpert C.A.;
RT   "The gal genes for the Leloir pathway of Lactobacillus casei 64H.";
RL   Appl. Environ. Microbiol. 64:2013-2019(1998).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: Galactose metabolism; third step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF005933; AAC19329.1; -.
DR   HSSP; P09147; 1XEL.
DR   InterPro; IPR001509; Epimerase_Dh.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF01370; Epimerase; 1.
DR   TIGRFAMs; TIGR01179; galE; 1.
KW   Isomerase; NAD; Galactose metabolism.
FT   NP_BIND      2     33       NAD (POTENTIAL).
SQ   SEQUENCE   331 AA;  36338 MW;  420DBDB3311EA7BF CRC64;

  Query Match            17.1%;  Score 72.5;  DB 1;  Length 331;
  Best Local Similarity  31.1%;  Pred. No. 0.53;
  Matches   19;  Conservative   13;  Mismatches   20;  Indels    9;  Gaps    2;

Qy         13 LLDLVRAHTMAV-LNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVT 71
              ::||  || :|:   | ||: :| : | :  |::|        |: |   | ||  :| |
Db        226 VVDLADAHILALKYLDAGNKSSAFNIGSAHGFSNL--------EILNAARKVTGQEIPAT 277

Qy         72 L 72
              :
Db        278 M 278


RESULT 14
PURL_LACCA
ID   PURL_LACCA     STANDARD;      PRT;   740 AA.
AC   P35852;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phosphoribosylformylglycinamidine synthase II (EC 6.3.5.3) (FGAM
DE   synthase II).
GN   PURL.
OS   Lactobacillus casei.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1582;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93012962; PubMed=1398079;
RA   Gu Z.-M., Martindale D.W., Lee B.H.;
RT   "Isolation and complete sequence of the purL gene encoding FGAM
RT   synthase II in Lactobacillus casei.";
RL   Gene 119:123-126(1992).
RN   [2]
RP   REVISIONS.
RX   MEDLINE=94040790; PubMed=8224889;
RA   Gu Z.-M., Martindale D.W., Lee B.H.;
RL   Gene 133:147-147(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
CC       ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC       (formamido)-N(1)-(5-phospho-D-ribosyl)acetamidine + L-glutamate.
CC   -!- PATHWAY: De novo purine biosynthesis; fourth step.
CC   -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURQ AND PURL.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M85265; AAC36947.1; -.
DR   PIR; JC1290; JC1290.
DR   InterPro; IPR000728; AIRS_related.
DR   Pfam; PF00586; AIRS; 2.
DR   Pfam; PF02769; AIRS_C; 2.
KW   Purine biosynthesis; Ligase; ATP-binding.
FT   NP_BIND     112    123       ATP (POTENTIAL).
SQ   SEQUENCE   740 AA;  79481 MW;  03238AB3CD1C8FF5 CRC64;

  Query Match            17.0%;  Score 72;  DB 1;  Length 740;
  Best Local Similarity  29.3%;  Pred. No. 1.6;
  Matches   17;  Conservative   10;  Mismatches   25;  Indels    6;  Gaps    1;

Qy          8 QQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATG 65
              | |  :|  :| | :   :|        || |   : | :|||: :| ::   |  | |
Db        618 QHQHFVLKAIREHLITAAHD------LSDGGLLVALAEMGFDAQLGAQINVTLPTAWG 669


RESULT 15
LYS2_YEAST
ID   LYS2_YEAST     STANDARD;      PRT;  1392 AA.
AC   P07702;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aminoadipate-semialdehyde dehydrogenase large subunit (EC 1.2.1.31)
DE   (Alpha-aminoadipate reductase) (Alpha-AR).
GN   LYS2 OR YBR115C OR YBR0910.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91192607; PubMed=2013406;
RA   Morris M.E., Jinks-Robertson S.;
RT   "Nucleotide sequence of the LYS2 gene of Saccharomyces cerevisiae:
```

```
RT   homology to Bacillus brevis tyrocidine synthetase 1.";
RL   Gene 98:141-145(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95208357; PubMed=7900426;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
RN   [3]
RP   SEQUENCE OF 1-150 AND 1209-1392 FROM N.A.
RX   MEDLINE=87106859; PubMed=3542721;
RA   Fleig U.N., Pridmore R.D., Philippsen P.;
RT   "Construction of LYS2 cartridges for use in genetic manipulations of
RT   Saccharomyces cerevisiae.";
RL   Gene 46:237-245(1986).
RN   [4]
RP   SEQUENCE OF 1083-1392 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [5]
RP   SEQUENCE OF 1-130 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94039074; PubMed=7916691;
RA   Schaaff-Gerstenschlager I., Mannhaupt G., Vetter I., Zimmermann F.K.,
RA   Feldmann H.;
RT   "TKL2, a second transketolase gene of Saccharomyces cerevisiae.
RT   Cloning, sequence and deletion analysis of the gene.";
RL   Eur. J. Biochem. 217:487-492(1993).
CC   -!- FUNCTION: CATALYZES THE ACTIVATION OF ALPHA-AMINOADIPATE BY
CC       ATP-DEPENDENT ADENYLATION AND THE REDUCTION OF ACTIVATED
CC       ALPHA-AMINOADIPATE BY NADPH.
CC   -!- CATALYTIC ACTIVITY: L-2-aminoadipate 6-semialdehyde + NADP(+) +
CC       H(2)O = L-2-aminoadipate + NADPH.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE
CC       (POTENTIAL).
CC   -!- PATHWAY: Lysine biosynthesis; sixth step.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36287; AAA34747.1; -.
DR   EMBL; X66247; CAA46975.1; -.
DR   EMBL; X78993; CAA55617.1; -.
DR   EMBL; Z35984; CAA85072.1; -.
DR   EMBL; X73532; CAA51938.1; -.
DR   PIR; JU0448; YGBYAD.
DR   HSSP; P14687; 1AMU.
DR   SGD; S0000319; LYS2.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Lysine biosynthesis; Oxidoreductase; NADP; Phosphopantetheine.
FT   DOMAIN      848    917       ACYL CARRIER (ACP).
FT   BINDING     880    880       PHOSPHOPANTETHEINE (POTENTIAL).

SQ   SEQUENCE   1392 AA;  155345 MW;  F0083A80BC6F7FB5 CRC64;

  Query Match             17.0%;  Score 72;  DB 1;  Length 1392;
  Best Local Similarity   28.1%;  Pred. No. 3.4;
  Matches   27;  Conservative   14;  Mismatches   37;  Indels   18;  Gaps    3;

Qy          6 PTQQQSLLLDLVRAHTMAVLNDD---GNERTASD-----------AGPSASFAHLGFDS 50
              || :|   |:||  :|::  :|      ||   |              |  ||  ||  |
Db        824 PTPKQ---LNLVAENTVSETDDSQFTNVEREVRDLWLSILPTKPASVSPDDSFFDLGGHS 880

Qy         51 VMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARL 86
                ::  ::   | |   : ||:  || : |  | || :
Db        881 ILATKMIFTLKKKLQVDLPLGTIFKYPTIKAFAAEI 916


Search completed: June 17, 2003, 13:02:32
Job time : 1.93786 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 1.92037 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_366_451
Perfect score: 424
Sequence: 1 LTALSPTQQQSLLLDLVRAH.....RLPVTLIFDHTTPAAVAARL 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	199	46.9	4613	2	T17409	polyketide synthas
2	195	46.0	1562	2	T17411	polyketide synthas
3	193	45.5	3413	2	T17467	rifamycin polyketi
4	189	44.6	2100	2	T03223	probable polyketid
5	184	43.4	1763	2	T17465	rifamycin polyketi
6	184	43.4	5069	2	T17464	rifamycin polyketi
7	183	43.2	4735	2	T17463	rifamycin polyketi
8	182	42.9	3739	2	T17410	polyketide synthas
9	181	42.7	1762	2	T03222	probable polyketid
10	180	42.5	6260	2	T30228	polyketide synthas
11	180	42.5	10223	2	T30225	polyketide synthas
12	179	42.2	1616	2	G70668	polyketide synthas
13	178	42.0	8563	2	T30226	polyketide synthas
14	176.5	41.6	7576	2	T17428	FK506 polyketide s
15	176	41.5	1728	2	T17466	rifamycin polyketi
16	174	41.0	2103	2	G86925	probable polyketid
17	173	40.8	3491	2	T43231	probable 6-deoxyer
18	169	39.9	2126	2	H70621	probable polyketid
19	168	39.6	1198	2	T28678	polyketide synthas
20	168	39.6	4151	2	G70944	probable polyketid
21	168	39.6	6420	2	T30283	polyketide synthas
22	166	39.2	3519	2	S43048	polyketide synthas
23	162.5	38.3	502	2	A70985	probable polyketid
24	161	38.0	3573	2	S23070	erythronolide synt
25	160.5	37.9	1017	2	B70985	probable polyketid
26	155	36.6	2124	2	T28658	polyketide synthas
27	150	35.4	1402	2	D70634	probable polyketid
28	138	32.5	3172	2	S22012	erythronolide synt
29	138	32.5	3178	2	S13595	6-deoxyerythronoli
30	137	32.3	1346	2	T17412	polyketide synthas
31	135	31.8	2723	2	T03221	probable polyketid
32	133	31.4	1937	2	T03224	probable polyketid
33	109	25.7	2201	2	S73014	polyketide synthas
34	104	24.5	707	2	P86925	probable acyl-CoA
35	104	24.5	2188	2	A70984	probable polyketid
36	102	24.1	705	2	A70669	probable acid-CoA
37	99	23.3	1538	2	E70874	probable ppsB prot
38	96.5	22.8	89	2	S18958	fix23-6 protein -
39	96	22.6	1540	2	H87203	polyketide synthas
40	94.5	22.3	130	2	S73018	hypothetical prote
41	91.5	21.6	622	2	A61197	6-methylsalicilic
42	90.5	21.3	1774	2	S13178	6-methylsalicylic
43	90	21.2	8243	2	T31307	type I fatty acid
44	88.5	20.9	15281	2	S41309	cyclosporin synthe
45	87	20.5	1004	2	H70673	probable mbtD prot

ALIGNMENTS

RESULT 1
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

Query Match 46.9%; Score 199; DB 2; Length 4613;
Best Local Similarity 48.8%; Pred. No. 1.1e-13;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :|||: ||||||||
Db       2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                ::||||:|| ||:||| || |: : |
Db       2541 TRATGLQLPATLVFDHPTPLALVSLL 2566
```

RESULT 2
T17411
polyketide synthase III - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17411
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17411
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1562 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C;Genetics:

```
A;Gene: pikAIII
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1403-1474/Domain: acyl carrier protein homology <ACP>

  Query Match             46.0%;  Score 195;  DB 2;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 8.5e-14;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 3
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

  Query Match             45.5%;  Score 193;  DB 2;  Length 3413;
  Best Local Similarity   47.7%;  Pred. No. 3.7e-13;
  Matches   41;  Conservative   14;  Mismatches   27;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:| :|::||||:||     ||   | |   :|     :|   ||||:  |||||||
Db       1598 LAGLAPAEQEALLLDVVRTQVALVLGHAGPEAVRAD----TAFKDTGFDSLTSVELRNRL 1653

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :|:||:|| ||:||: || |:|  |
Db       1654 REASGLKLPATLVFDYPTPVALARYL 1679


RESULT 4
T03223
probable polyketide synthase module 3 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03223
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2100 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38063.1; PID:g2624950
A;Experimental source: ATCC 29253
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;566-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1952-2027/Domain: acyl carrier protein homology <ACP>

  Query Match             44.6%;  Score 189;  DB 2;  Length 2100;
  Best Local Similarity   45.3%;  Pred. No. 6e-13;
  Matches   39;  Conservative   17;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              : |::| ::::::|:||||    ||   | |  :    |: :|   ||||:  || |||:
Db       1942 VAAVAPEERRAVVLELVRAQAATVLGHSGAEHVS----PTQAFKDSGFDSLTAVEFRNRV 1997

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| |||||| ||  :|  |
Db       1998 TTATGLRLPATLIFDHPTPNTLADHL 2023


RESULT 5
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>

  Query Match             43.4%;  Score 184;  DB 2;  Length 1763;
  Best Local Similarity   46.5%;  Pred. No. 1.8e-12;
  Matches   40;  Conservative   12;  Mismatches   30;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:  :|::|||||||     ||    :    :|    |:|   ||||:  |||||||
Db       1605 LAGLTVAEQEALLLDLVRGQVAVVLGHADSSGVRAD----AAFKDAGFDSLTSVELRNRL 1660

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : |||:|| ||:|||  | |:|  |
Db       1661 RETTGLKLPATLVFDHPNPLALARHL 1686


RESULT 6
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             43.4%;  Score 184;  DB 2;  Length 5069;
  Best Local Similarity   48.8%;  Pred. No. 6.5e-12;
  Matches   42;  Conservative    9;  Mismatches   31;  Indels    4;  Gaps    1;
```

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:|  | :||||||||   |||        ||      :|   ||||: .||||||:
Db       3228 LAGLAPADQLTLLLDLVRAQVAAVL----GHADASAVRVDTAFKDAGFDSLTAVELRNRM 3283

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                ||||:|| ||:||:  | |:|  |
Db       3284 RTATGLKLPATLVFDYPNPQALARHL 3309


RESULT 7
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

  Query Match             43.2%;  Score 183;  DB 2;  Length 4735;
  Best Local Similarity   47.7%;  Pred. No. 7.8e-12;
  Matches   41;  Conservative    8;  Mismatches   33;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | |:   :|  :|| ||| |   ||   |    |   || :|  :||||:  | ||| |
Db       4568 LRAVPDAEQNRILLKLVRGHASTVLGHSG----AEGIGPRQAFQEVGFDSLAAVNLRNSL 4623

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                |||||| |||||: || |:   |
Db       4624 HAATGLRLPATLIFDYPTPEALVGYL 4649


RESULT 8
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

  Query Match             42.9%;  Score 182;  DB 2;  Length 3739;
  Best Local Similarity   47.7%;  Pred. No. 7.6e-12;
  Matches   41;  Conservative    9;  Mismatches   32;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | |::|  : : | |||| |   ||       | |      :|   ||||: |||||||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVL----GHGTPSRVDLERAFRDTGFDSLTAVELRNRL 3615

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                : ||||||| ||:||| ||  :|  |
Db       3616 NAATGLRLPATLVFDHPTPGELAGHL 3641


RESULT 9
T03222
probable polyketide synthase module 2 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Nov-2000
C;Accession: T03222
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03222
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1762 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38062.1; PID:g2624949
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;550-822/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1598-1669/Domain: acyl carrier protein homology <ACP>

  Query Match             42.7%;  Score 181;  DB 2;  Length 1762;
  Best Local Similarity   46.5%;  Pred. No. 4e-12;
  Matches   40;  Conservative   11;  Mismatches   31;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              : || | ::   :|||||  :  ||     :|       |  :|   ||||:  || |||:
Db       1588 IAALMPEERDRAVLDLVRQQSARVLGHSDADR----IDPGQAFKDTGFDSLTAVEFRNRV 1643

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| |||||| || |:|  |
Db       1644 TTATGLRLPATLIFDHPTPNALADHL 1669


RESULT 10
T30228
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30228
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30228
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6260 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987102; PIDN:CAA60462.1
C;Genetics:
A;Gene: rapC
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1536-1607/Domain: acyl carrier protein homology <ACP1>
F;1651-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3123-3194/Domain: acyl carrier protein homology <ACP2>
F;3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3725-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5114-5185/Domain: acyl carrier protein homology <ACP3>
F;5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6135-6206/Domain: acyl carrier protein homology <ACP4>
```

```
  Query Match             42.5%;  Score 180;  DB 2;  Length 6260;
  Best Local Similarity   48.8%;  Pred. No. 2.4e-11;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | ||:| :::  || |||     ||        ||    :|:|  || ||:  ||||| |
Db       3113 LAALAPAEREDALLKLVRDSAALVL----GHADASTIPAAAAFKDLGIDSLTAVELRNSL 3168

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :|||||||| | :||: ||| :| ||
Db       3169 AKATGLRLPNTTVFDYPTPAILATRL 3194


RESULT 11
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match             42.5%;  Score 180;  DB 2;  Length 10223;
  Best Local Similarity   48.8%;  Pred. No. 4.3e-11;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:| :::  || ||      ||        ||    :|:|  || ||:  ||||| |
Db       5150 LARLAPVEREKALLKLVCDGAATVL----GHADASTIPATAAFKDLGIDSLTAVELRNSL 5205

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :|||||||| ||:||: || |:||||
Db       5206 TKATGLRLPATLVFDYPTPTALAARL 5231


RESULT 12
G70668
polyketide synthase pks1 - Mycobacterium tuberculosis (strain H37RV)
N;Alternate names: polyketide synthase pks002c
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70668; S73075
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70668
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1616 <COL>
A;Cross-references: GB:Z83858; GB:AL123456; NID:g3261675; PIDN:CAB06103.1; PID:g1781167
A;Experimental source: strain H37Rv
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A;Description: Mycobacterium tuberculosis cosmid tbc2.
A;Reference number: S73053
A;Accession: S73075
A;Molecule type: DNA
A;Residues: 'MWGARWPVARCLSTEL',48-917,'RSEARSRRGCGMSGRAAPPAPPIGPPIGIP','PMPPIIPVEPAGPVFI
A;Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50928.1; PID:g560507
C;Genetics:
A;Gene: pks1
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase pks1; acyl carrie
ology
C;Keywords: carrier protein
F;84-362/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1228-1410/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1514-1585/Domain: acyl carrier protein homology <ACP1>

  Query Match             42.2%;  Score 179;  DB 2;  Length 1616;
  Best Local Similarity   50.6%;  Pred. No. 6.1e-12;
  Matches   42;  Conservative    6;  Mismatches   31;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:  :|  ||: ||     |||          :| |  |   |  |||||:  |||||||
Db       1504 LHGLAADEQLELLVGLVCLQAAAVL----GRPSAEDVDPDTEFGDLGFDSLTAVELRNRL 1559

Qy         61 SKATGLRLPVTLIFDHTTPAAVA 83
              |||| || |:|||| || |||
Db       1560 KTATGLTLPPTVIFDHPTPTAVA 1582


RESULT 13
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staur
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             42.0%;  Score 178;  DB 2;  Length 8563;
  Best Local Similarity   48.8%;  Pred. No. 5.9e-11;
```

```
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:| :::  || ||      ||      ||    : :|  || ||:  ||||| |
Db       3189 LARLAPVEREKALLKLVCDGAATVL----GHADASTIPATGAFRDLGVDSLTAVELRNGL 3244

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :|||||||| ||:||: ||||:||||
Db       3245 AKATGLRLPATLVFDYPTPAALAARL 3270


RESULT 14
T17428
FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: T17428
R;Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressan
A;Reference number: Z18779; MUID:98451508; PMID:9780228
A;Accession: T17428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7576 <MOT>
A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A;Experimental source: strain MA6548
C;Genetics:
A;Gene: fkbB
C;Function:
A;Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-500/Domain: acetate-CoA ligase homology <ACL>
F;1095-1166/Domain: acyl carrier protein homology <ACP1>
F;1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;2680-2751/Domain: acyl carrier protein homology <ACP2>
F;2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;4320-4391/Domain: acyl carrier protein homology <ACP3>
F;4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5903-5974/Domain: acyl carrier protein homology <ACP4>
F;6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;7450-7521/Domain: acyl carrier protein homology <ACP5>

  Query Match             41.6%;  Score 176.5;  DB 2;  Length 7576;
  Best Local Similarity   48.8%;  Pred. No. 7.6e-11;
  Matches   41;  Conservative   11;  Mismatches   27;  Indels    5;  Gaps    2;

Qy          3 ALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSK 62
              | :||  :  :|:|||     ||      | |    |: :|  :||||: ||||||||:
Db       4313 AAAPTAARD-MLELVRTSAALVL----GHRDAHAVEPTRAFKEVGFDSLTGVELRNRLAD 4367

Qy         63 ATGLRLPVTLIFDHTTPAAVAARL 86
              |||| || ||:||| |  |:|| |
Db       4368 ATGLTLPATLVFDHPTAQALAAHL 4391


RESULT 15
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17466
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17466
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1728 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1573-1644/Domain: acyl carrier protein homology <ACP>

  Query Match             41.5%;  Score 176;  DB 2;  Length 1728;
  Best Local Similarity   44.2%;  Pred. No. 1.5e-11;
  Matches   38;  Conservative   14;  Mismatches   30;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |   :|::||:||||    |||   | :   :|     :|   ||||:  |:|||||
Db       1563 LAGLGAPEQEALLVDLVRGQVAAVLGHAGPDAVRAD----TAFKDAGFDSLTSVDLRNRL 1618

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::|||:|| || ||: ||  :|  |
Db       1619 RESTGLKLPATLAFDYPTPLVLARHL 1644


Search completed: June 17, 2003, 13:12:19
Job time : 1.92037 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 2.67959 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_366_451
Perfect score: 424
Sequence: 1 LTALSPTQQQSLLLDLVRAH.....RLPVTLIFDHTTPAAVAARL 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	199	46.9	4551	9	US-09-793-708-1	Sequence 1, Appli
2	199	46.9	4613	9	US-09-860-846-31	Sequence 31, Appl
3	199	46.9	4613	9	US-09-988-384B-31	Sequence 31, Appl
4	199	46.9	4613	9	US-09-836-821-31	Sequence 31, Appl
5	199	46.9	4613	10	US-09-861-289-31	Sequence 31, Appl
6	199	46.9	11877	9	US-09-860-846-6	Sequence 6, Appli
7	199	46.9	11877	9	US-09-836-821-6	Sequence 6, Appli
8	199	46.9	11877	10	US-09-861-289-6	Sequence 6, Appli
9	199	46.9	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	196	46.2	4150	9	US-09-808-880-2	Sequence 2, Appli
11	195	46.0	1562	9	US-09-860-846-35	Sequence 35, Appl
12	195	46.0	1562	9	US-09-988-384B-35	Sequence 35, Appl
13	195	46.0	1562	9	US-09-836-821-35	Sequence 35, Appl
14	195	46.0	1562	9	US-09-793-708-3	Sequence 3, Appli
15	195	46.0	1562	10	US-09-861-289-35	Sequence 35, Appl
16	183	43.2	5215	9	US-09-860-846-2	Sequence 2, Appli
17	183	43.2	5215	9	US-09-988-384B-2	Sequence 2, Appli
18	183	43.2	5215	9	US-09-836-821-2	Sequence 2, Appli
19	183	43.2	5215	10	US-09-861-289-2	Sequence 2, Appli
20	182	42.9	3739	9	US-09-860-846-33	Sequence 33, Appl
21	182	42.9	3739	9	US-09-988-384B-33	Sequence 33, Appl
22	182	42.9	3739	9	US-09-836-821-33	Sequence 33, Appl
23	182	42.9	3739	9	US-09-793-708-2	Sequence 2, Appli
24	182	42.9	3739	10	US-09-861-289-33	Sequence 33, Appl
25	179	42.2	1616	9	US-09-712-363-262	Sequence 262, App
26	166.5	39.3	3816	9	US-09-808-880-3	Sequence 3, Appli
27	166	39.2	3519	9	US-09-808-880-4	Sequence 4, Appli
28	150	35.4	1402	9	US-09-712-363-166	Sequence 166, App
29	139	32.8	1346	9	US-09-793-708-4	Sequence 4, Appli
30	137	32.3	1346	9	US-09-860-846-37	Sequence 37, Appl
31	137	32.3	1346	9	US-09-988-384B-37	Sequence 37, Appl
32	137	32.3	1346	9	US-09-836-821-37	Sequence 37, Appl
33	137	32.3	1346	10	US-09-861-289-37	Sequence 37, Appl
34	130	30.7	3798	9	US-10-014-717-6	Sequence 6, Appli
35	117.5	27.7	1832	9	US-10-014-717-4	Sequence 4, Appli
36	115.5	27.2	7257	9	US-10-014-717-5	Sequence 5, Appli
37	115	27.1	352	10	US-09-073-009-14	Sequence 14, Appl
38	115	27.1	352	10	US-09-023-588-14	Sequence 14, Appl
39	115	27.1	352	10	US-09-793-306-14	Sequence 14, Appl
40	103.5	24.4	2439	9	US-10-014-717-7	Sequence 7, Appli
41	92.5	21.8	1421	9	US-10-014-717-2	Sequence 2, Appli
42	83	19.6	4999	9	US-09-976-059-15	Sequence 15, Appl
43	82	19.3	4999	9	US-09-976-059-14	Sequence 14, Appl
44	79	18.6	1812	10	US-09-775-938A-38	Sequence 38, Appl
45	78.5	18.5	90	9	US-09-976-059-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-793-708-1

Query Match 46.9%; Score 199; DB 9; Length 4551;
Best Local Similarity 48.8%; Pred. No. 9.6e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| |||| |||      |: |     :|  :|||: ||||||||
Db       2423 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2478
```

```
Qy          61 SKATGLRLPVTLIFDHTTPAAVAARL 86
               ::||||:|| ||:||| || |: : |
Db        2479 TRATGLQLPATLVFDHPTPLALVSLL 2504
```

RESULT 2
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-31

Query Match 46.9%; Score 199; DB 9; Length 4613;
Best Local Similarity 48.8%; Pred. No. 9.8e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy           1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
               | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db        2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy          61 SKATGLRLPVTLIFDHTTPAAVAARL 86
               ::||||:|| ||:||| || |: : |
Db        2541 TRATGLQLPATLVFDHPTPLALVSLL 2566
```

RESULT 3
US-09-988-384B-31
; Sequence 31, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-31

Query Match 46.9%; Score 199; DB 9; Length 4613;
Best Local Similarity 48.8%; Pred. No. 9.8e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy           1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
               | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db        2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy          61 SKATGLRLPVTLIFDHTTPAAVAARL 86
               ::||||:|| ||:||| || |: : |
Db        2541 TRATGLQLPATLVFDHPTPLALVSLL 2566
```

RESULT 4
US-09-836-821-31
; Sequence 31, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-31

Query Match 46.9%; Score 199; DB 9; Length 4613;
Best Local Similarity 48.8%; Pred. No. 9.8e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy           1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
               | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db        2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy          61 SKATGLRLPVTLIFDHTTPAAVAARL 86
               ::||||:|| ||:||| || |: : |
Db        2541 TRATGLQLPATLVFDHPTPLALVSLL 2566
```

RESULT 5
US-09-861-289-31
; Sequence 31, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-31

Query Match 46.9%; Score 199; DB 10; Length 4613;
Best Local Similarity 48.8%; Pred. No. 9.8e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy           1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
               | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db        2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy          61 SKATGLRLPVTLIFDHTTPAAVAARL 86
```

```
              ::||||:|| ||:||| || |: : |
Db       2541 TRATGLQLPATLVFDHPTPLALVSLL 2566


RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match             46.9%;  Score 199;  DB 9;  Length 11877;
  Best Local Similarity   48.8%;  Pred. No. 3.5e-15;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db       2821 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2876

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2877 TRATGLQLPATLVFDHPTPLALVSLL 2902


RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-6

  Query Match             46.9%;  Score 199;  DB 9;  Length 11877;
  Best Local Similarity   48.8%;  Pred. No. 3.5e-15;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db       2821 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2876

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2877 TRATGLQLPATLVFDHPTPLALVSLL 2902


RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             46.9%;  Score 199;  DB 10;  Length 11877;
  Best Local Similarity   48.8%;  Pred. No. 3.5e-15;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db       2821 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2876

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2877 TRATGLQLPATLVFDHPTPLALVSLL 2902


RESULT 9
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             46.9%;  Score 199;  DB 9;  Length 12199;
  Best Local Similarity   48.8%;  Pred. No. 3.6e-15;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db       3143 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 3198

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
```

```
Db       3199 TRATGLQLPATLVFDHPTPLALVSLL 3224


RESULT 10
US-09-808-880-2
; Sequence 2, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-2

  Query Match           46.2%;  Score 196;  DB 9;  Length 4150;
  Best Local Similarity   50.0%;  Pred. No. 2.1e-15;
  Matches   43;  Conservative   10;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:  :|: ||: |||     ||   | |  |    | :|  |||||:  |||||||
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVA----PDRAFKDLGFDSLTSVELRNRL 2496

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : |||||||| :||:  |||:|  |
Db       2497 NTATGLRLPVTAVFDYARPAALAGHL 2522


RESULT 11
US-09-860-846-35
; Sequence 35, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-35

  Query Match           46.0%;  Score 195;  DB 9;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 7.4e-16;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| | -|||     :| |    | :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 12
US-09-988-384B-35
; Sequence 35, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-35

  Query Match           46.0%;  Score 195;  DB 9;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 7.4e-16;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| | |||     :| |    | :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 13
US-09-836-821-35
; Sequence 35, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-35

  Query Match           46.0%;  Score 195;  DB 9;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 7.4e-16;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;
```

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 14
US-09-793-708-3
; Sequence 3, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-3

  Query Match             46.0%;  Score 195;  DB 9;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 7.4e-16;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 15
US-09-861-289-35
; Sequence 35, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-35

  Query Match             46.0%;  Score 195;  DB 10;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 7.4e-16;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


Search completed: June 17, 2003, 13:46:40
Job time : 3.67959 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 1.5631 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_366_451
Perfect score: 424
Sequence: 1 LTALSPTQQQSLLLDLVRAH.....RLPVTLIFDHTTPAAVAARL 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	199	46.9	4551	3	US-09-320-878-1	Sequence 1, Appli
2	199	46.9	4613	4	US-09-105-537-31	Sequence 31, Appl
3	199	46.9	4928	4	US-09-036-987A-5	Sequence 5, Appli
4	199	46.9	4928	4	US-09-370-700-5	Sequence 5, Appli
5	199	46.9	11877	4	US-09-105-537-6	Sequence 6, Appli
6	196	46.2	4150	4	US-09-428-517-2	Sequence 2, Appli
7	195	46.0	1562	3	US-09-320-878-3	Sequence 3, Appli
8	195	46.0	1562	4	US-09-105-537-35	Sequence 35, Appl
9	192	45.3	2152	4	US-09-036-987A-3	Sequence 3, Appli
10	192	45.3	2152	4	US-09-370-700-3	Sequence 3, Appli
11	190	44.8	3724	2	US-08-804-227C-10	Sequence 10, Appl
12	190	44.8	3724	2	US-08-804-198-4	Sequence 4, Appli
13	190	44.8	3729	2	US-08-804-227C-4	Sequence 4, Appli
14	183	43.2	5215	4	US-09-105-537-2	Sequence 2, Appli
15	182	42.9	3739	3	US-09-320-878-2	Sequence 2, Appli
16	182	42.9	3739	4	US-09-105-537-33	Sequence 33, Appl
17	181	42.7	3170	4	US-09-036-987A-4	Sequence 4, Appli
18	181	42.7	3170	4	US-09-370-700-4	Sequence 4, Appli
19	180	42.5	185	4	US-09-154-083-14	Sequence 14, Appl
20	178	42.0	1611	2	US-08-804-227C-5	Sequence 5, Appli
21	177	41.7	502	4	US-09-413-814-106	Sequence 106, App
22	176.5	41.6	2595	4	US-09-036-987A-2	Sequence 2, Appli
23	176.5	41.6	2595	4	US-09-370-700-2	Sequence 2, Appli
24	176	41.5	1996	2	US-08-804-227C-9	Sequence 9, Appli
25	176	41.5	1996	2	US-08-804-198-3	Sequence 3, Appli
26	174	41.0	3567	2	US-07-642-734C-4	Sequence 4, Appli
27	174	41.0	3567	3	US-08-439-009A-4	Sequence 4, Appli
28	173	40.8	3491	2	US-07-642-734C-2	Sequence 2, Appli
29	173	40.8	3491	3	US-08-439-009A-2	Sequence 2, Appli
30	171	40.3	5588	4	US-09-036-987A-6	Sequence 6, Appli
31	171	40.3	5588	4	US-09-370-700-6	Sequence 6, Appli
32	168	39.6	4472	2	US-08-804-227C-2	Sequence 2, Appli
33	168	39.6	4545	2	US-08-804-227C-14	Sequence 14, Appl
34	167	39.4	4550	2	US-08-804-227C-8	Sequence 8, Appli
35	167	39.4	4550	2	US-08-804-198-2	Sequence 2, Appli
36	166.5	39.3	3816	4	US-09-428-517-3	Sequence 3, Appli
37	166	39.2	3519	4	US-09-428-517-4	Sequence 4, Appli
38	164	38.7	1864	2	US-08-804-227C-3	Sequence 3, Appli
39	162	38.2	1841	2	US-08-804-227C-6	Sequence 6, Appli
40	159	37.5	1580	2	US-08-804-227C-11	Sequence 11, Appl
41	159	37.5	1580	2	US-08-804-198-5	Sequence 5, Appli
42	155	36.6	1891	2	US-08-804-227C-12	Sequence 12, Appl
43	155	36.6	1891	2	US-08-804-198-6	Sequence 6, Appli
44	152	35.8	5087	4	US-09-144-085-1	Sequence 1, Appli
45	147	34.7	6095	4	US-09-144-085-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-1

Query Match 46.9%; Score 199; DB 3; Length 4551;
Best Local Similarity 48.8%; Pred. No. 7.7e-17;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| |||| |||      |: |     :|  :||||: ||||||||
Db       2423 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2478

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                ::||||:|| ||:||| || |: : |
Db       2479 TRATGLQLPATLVFDHPTPLALVSLL 2504
```

RESULT 2
US-09-105-537-31
; Sequence 31, Application US/09105537A

```
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             46.9%;  Score 199;  DB 4;  Length 4613;
  Best Local Similarity   48.8%;  Pred. No. 7.8e-17;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |: |     :|  :||||: ||||||||
Db       2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2541 TRATGLQLPATLVFDHPTPLALVSLL 2566


RESULT 3
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4928 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-5

  Query Match             46.9%;  Score 199;  DB 4;  Length 4928;
  Best Local Similarity   50.0%;  Pred. No. 8.6e-17;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | || | ::  |||::| | | : |  |    :|    |  :|| |||||:  |||||||
Db       4762 LAALPPAERGHLLLEVVLAETASTLGHD----SAEAVQPDRTFAELGFDSLTAVELRNRL 4817

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :  |||||| ||:||| || |:: :|
Db       4818 NAVTGLRLPPTLVFDHPTPLALSEQL 4843


RESULT 4
US-09-370-700-5
; Sequence 5, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 4928
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-5

  Query Match             46.9%;  Score 199;  DB 4;  Length 4928;
  Best Local Similarity   50.0%;  Pred. No. 8.6e-17;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | || | ::  |||::| | | : |  |    :|    |  :|| |||||:  |||||||
Db       4762 LAALPPAERGHLLLEVVLAETASTLGHD----SAEAVQPDRTFAELGFDSLTAVELRNRL 4817

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :  |||||| ||:||| || |:: :|
Db       4818 NAVTGLRLPPTLVFDHPTPLALSEQL 4843


RESULT 5
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
```

```
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             46.9%;  Score 199;  DB 4;  Length 11877;
  Best Local Similarity   48.8%;  Pred. No. 3e-16;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||  |||     |:  |     :|  :||||: ||||||||
Db       2821 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2876

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2877 TRATGLQLPATLVFDHPTPLALVSLL 2902


RESULT 6
US-09-428-517-2
; Sequence 2, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-2

  Query Match             46.2%;  Score 196;  DB 4;  Length 4150;
  Best Local Similarity   50.0%;  Pred. No. 1.7e-16;
  Matches   43;  Conservative   10;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:  :|: ||: |||     ||   | |  |    | :|  |||||:  |||||||
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVA----PDRAFKDLGFDSLTSVELRNRL 2496

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : |||||||| :||:  |||:|  |
Db       2497 NTATGLRLPVTAVFDYARPAALAGHL 2522


RESULT 7
US-09-320-878-3
; Sequence 3, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-3

  Query Match             46.0%;  Score 195;  DB 3;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 5.8e-17;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 8
US-09-105-537-35
; Sequence 35, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-35

  Query Match             46.0%;  Score 195;  DB 4;  Length 1562;
  Best Local Similarity   48.8%;  Pred. No. 5.8e-17;
  Matches   42;  Conservative   11;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              :|||   :::  || ||| |  |||     :| |    |  :|  |||||:  |:|||:|
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474


RESULT 9
US-09-036-987A-3
; Sequence 3, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
```

```
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2152 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-3

  Query Match             45.3%;  Score 192;  DB 4;  Length 2152;
  Best Local Similarity   50.0%;  Pred. No. 2.3e-16;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |    :|  :|||||| |  |||    |::   :|     :|  |||||:  ||||||:
Db       1989 LGRLPDAEQHRILLDLVRMHVAAVLGFAGSQEITAD----GTFKVLGFDSLTVVELRNRI 2044

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| ||:|:: || |:|| |
Db       2045 NGATGLRLPATLVFNYPTPDALAAHL 2070


RESULT 10
US-09-370-700-3
; Sequence 3, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2152
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-3

  Query Match             45.3%;  Score 192;  DB 4;  Length 2152;
  Best Local Similarity   50.0%;  Pred. No. 2.3e-16;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |    :|  :|||||| |  |||    |::   :|     :|  |||||:  ||||||:
Db       1989 LGRLPDAEQHRILLDLVRMHVAAVLGFAGSQEITAD----GTFKVLGFDSLTVVELRNRI 2044

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| ||:|:: || |:|| |
Db       2045 NGATGLRLPATLVFNYPTPDALAAHL 2070


RESULT 11
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-10

  Query Match             44.8%;  Score 190;  DB 2;  Length 3724;
  Best Local Similarity   47.7%;  Pred. No. 9.2e-16;
  Matches   41;  Conservative   17;  Mismatches   24;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              ||| |  :: : || |||||  |||      :  :|    : :|  |||||:  :|||:||
Db       1440 LTAQSGHKRHATLLRLVRAHAAAVLGQSSGDAVSS----ARAFRDLGFDSLTALELRDRL 1495

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              | :|||:|| :|:|||::|||:|  |
Db       1496 STSTGLKLPTSLVFDHSSPAALARHL 1521
```

```
RESULT 12
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-4

  Query Match             44.8%;  Score 190;  DB 2;  Length 3724;
  Best Local Similarity   47.7%;  Pred. No. 9.2e-16;
  Matches   41;  Conservative   17;  Mismatches   24;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              ||| |  :: : || |||||  |||     :  :|    : :|  |||||:  :|||:||
Db       1440 LTAQSGHKRHATLLRLVRAHAAAVLGQSSGDAVSS----ARAFRDLGFDSLTALELRDRL 1495

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              | :|||:|| :|:|||::|||:|  |
Db       1496 STSTGLKLPTSLVFDHSSPAALARHL 1521


RESULT 13
US-08-804-227C-4
; Sequence 4, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3729 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-4

  Query Match             44.8%;  Score 190;  DB 2;  Length 3729;
  Best Local Similarity   47.7%;  Pred. No. 9.2e-16;
  Matches   41;  Conservative   12;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              || |:  :  :|:|  :|||  |||    ::    |     :|  |||||:  ||:||||
Db       3547 LTGLTAQEGHALVLAEIRAHAAAVLGHGSDDSIPED----RAFKDLGFDSLTAVEMRNRL 3602

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              | ||||||| ||:||| ||  :|  |
Db       3603 SAATGLRLPATLVFDHPTPGELAGHL 3628


RESULT 14
US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-2

  Query Match             43.2%;  Score 183;  DB 4;  Length 5215;
  Best Local Similarity   50.0%;  Pred. No. 1.3e-14;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    2;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | || :::|  : ||      ||   |:  :|:  |   ||  |||||: ||||||||
Db       2713 LAAQSPEERRSTAVTLVTGVVADVL---GHADSAA-VGAERSFKDLGFDSLAGVELRNRL 2768

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| | :||| :|||:|: |
Db       2769 NAATGLRLPATTVFDHPSPAALASHL 2794


RESULT 15
US-09-320-878-2
```

; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2

Query Match 42.9%; Score 182; DB 3; Length 3739;
Best Local Similarity 47.7%; Pred. No. 1.1e-14;
Matches 41; Conservative 9; Mismatches 32; Indels 4; Gaps 1;

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | |::|  : : | |||| |   ||       | |      :|   ||||:  |||||||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVL----GHGTPSRVDLERAFRDTGFDSLTAVELRNRL 3615

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||| ||:||| ||  :|  |
Db       3616 NAATGLRLPATLVFDHPTPGELAGHL 3641
```

Search completed: June 17, 2003, 13:09:25
Job time : 2.5631 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 4.21291 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_366_451
Perfect score: 424
Sequence: 1 LTALSPTQQQSLLLDLVRAH.....RLPVTLIFDHTTPAAVAARL 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	424	100.0	3972	21	AAB23749	S. avermitilis ave
2	424	100.0	3972	22	AAG65264	Streptomyces averm
3	424	100.0	3972	22	AAG65268	Streptomyces averm
4	204	48.1	5435	22	AAE10145	Streptomyces nours
5	199.5	47.1	47	23	AAM47200	Modular enzyme sys
6	199	46.9	4551	21	AAB18637	Amino acid sequenc
7	199	46.9	4551	21	AAY67201	Narbonolide syntha
8	199	46.9	4613	21	AAY77192	S. venezuelae macr
9	199	46.9	4613	21	AAY77200	S. venezuelae pik
10	199	46.9	4924	22	AAB70968	S. spinosa protein
11	199	46.9	4928	20	AAY39300	SpnD a polyketide
12	199	46.9	12199	21	AAY77180	S. venezuelae pik
13	196	46.2	4150	21	AAY92707	S. antibioticus 8,
14	195	46.0	1560	21	AAY67203	Narbonolide syntha
15	195	46.0	1562	21	AAB18639	Amino acid sequenc
16	195	46.0	1562	21	AAY77194	S. venezuelae macr
17	193.5	45.6	1561	21	AAY77202	S. venezuelae pik
18	193	45.5	1891	19	AAW52844	Amycolatopsis medi
19	193	45.5	3413	19	AAW52849	A. mediterranei ri
20	193	45.5	4881	21	AAB23751	S. avermitilis ave
21	193	45.5	4881	22	AAG65266	Streptomyces averm
22	192	45.3	2152	20	AAY39298	SpnB a polyketide
23	192	45.3	2152	22	AAB70966	S. spinosa protein
24	190	44.8	3724	18	AAW23718	Platenolide syntha
25	190	44.8	3724	18	AAW22608	Platenolide syntha
26	190	44.8	3729	18	AAW22603	Tylactone synthase
27	184.5	43.5	11096	22	AAE10129	Streptomyces nours
28	184	43.4	1721	19	AAW52847	A. mediterranei ri
29	184	43.4	5069	19	AAW52846	A. mediterranei ri
30	183	43.2	4572	19	AAW52845	A. mediterranei ri
31	183	43.2	4630	21	AAY77177	S. venezuelae vep
32	182	42.9	3739	21	AAB18638	Amino acid sequenc
33	182	42.9	3739	21	AAY77193	S. venezuelae macr
34	182	42.9	3739	21	AAY77201	S. venezuelae pik
35	182	42.9	3739	21	AAY67202	Narbonolide syntha
36	181	42.7	3170	20	AAY39299	SpnC a polyketide
37	181	42.7	3170	22	AAB70967	S. spinosa protein
38	180	42.5	186	22	AAB19902	Polyketide synthas
39	179	42.2	1366	22	AAE10127	Streptomyces nours
40	179	42.2	1615	22	AAG81211	Mycobacterium tube
41	179	42.2	6797	22	AAB31558	Pimaricin biosynth
42	178	42.0	1611	18	AAW22604	Tylactone synthase
43	177	41.7	3562	22	AAB82213	Polyketide synthas
44	177	41.7	4630	18	AAW19629	Streptomyces venez
45	177	41.7	6239	21	AAB23750	S. avermitilis ave

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 424; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 4.6e-46;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 366 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 425

Qy 61 SKATGLRLPVTLIFDHTTPAAVAARL 86
||||||||||||||||||||||||||
Db 426 SKATGLRLPVTLIFDHTTPAAVAARL 451

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 424; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 4.6e-46;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 366 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 425

Qy 61 SKATGLRLPVTLIFDHTTPAAVAARL 86
||||||||||||||||||||||||||
Db 426 SKATGLRLPVTLIFDHTTPAAVAARL 451

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 424; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 4.6e-46;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 425

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |||||||||||||||||||||||||
Db        426 SKATGLRLPVTLIFDHTTPAAVAARL 451
```

RESULT 4
AAE10145
ID AAE10145 standard; Protein; 5435 AA.
XX
AC AAE10145;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin gene, NysJ.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic.
XX
OS Streptomyces noursei.
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR N-PSDB; AAD17186.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 15; Page 260-263; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a PKS encoding Streptomyces noursei nystatin gene, NysJ.
XX
SQ Sequence 5435 AA;

```
  Query Match          48.1%;  Score 204;  DB 22;  Length 5435;
  Best Local Similarity   55.2%;  Pred. No. 9.8e-17;
  Matches   48;  Conservative    6;  Mismatches   27;  Indels    6;  Gaps    2;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPS-ASFAHLGFDSVMGVELRNR 59
              | ||   |:  ||||||  | |||   |     :|| |:   |  |||||:| |||| |
Db       5265 LAALDAEQRHQALLDLVRTETAAVLGHSG-----ADAVPAERDFNRLGFDSLMAVELRTR 5319

Qy         60 LSKATGLRLPVTLIFDHTTPAAVAARL 86
              |: ||| ||| ||:||| || |||  |
Db       5320 LATATGARLPATLVFDHPTPDAVARHL 5346
```

RESULT 5
AAM47200
ID AAM47200 standard; Peptide; 47 AA.
XX
AC AAM47200;
XX
DT 12-FEB-2002 (first entry)
XX
DE Modular enzyme system related ACP-domain N-terminal peptide AVEAI_1.
XX
KW Modular enzyme system; cyclic gene synthesis; repetitive coding sequence;
KW antibiotic; non-ribosomal peptide synthetase; NRPS; PKS;
KW polyketide synthase; actinomycin biosynthase.
XX
OS Streptomyces avermitilis.
XX
PN WO200181564-A2.
XX
PD 01-NOV-2001.
XX
PF 25-APR-2001; 2001WO-DE01578.
XX
PR 26-APR-2000; 2000DE-1021267.
XX
PA (ACTI-) ACTINODRUG PHARM GMBH.
XX
PI Schauwecker F;
XX
DR WPI; 2002-049276/06.
XX
PT Preparing DNA encoding modular protein for e.g. producing new enzymes
PT for synthesis of polyketide antibiotics, comprises cyclic integration
PT of fragments into a vector -
XX
PS Example 1; Fig 9; 83pp; German.
XX
CC The present invention relates to the preparation of DNA, in a circular
CC vector, that encodes one or more segments of a modular polypeptide. DNA
CC or DNA libraries produced this way are used to produce modular
CC polypeptides, particularly enzymes, which can be used to act on
CC substrates to produce compounds for therapeutic testing. Enzymes of
CC particular interest are those involved in non-ribosomal peptide synthesis
CC or polyketide synthesis, and compounds for testing are particularly
CC macrolide antibiotics, including penicillins, vancomycins or
CC erythromycins, but may also be modular receptors. The present sequence is
CC a peptide used in the exemplification of the invention.
XX
SQ Sequence 47 AA;

```
  Query Match          47.1%;  Score 199.5;  DB 23;  Length 47;
  Best Local Similarity   89.6%;  Pred. No. 6e-19;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    5;  Gaps    1;

Qy         39 PSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARL 86
              ||||||||     ||||||||||||||||||||||||||||||||||
Db          1 PSASFAHL----MGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARL 43
```

RESULT 6
AAB18637
ID AAB18637 standard; Protein; 4551 AA.
XX
AC AAB18637;
XX
DT 22-JAN-2001 (first entry)
XX
DE Amino acid sequence of narbonolide synthase subunit 1 (PICAI).
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;

KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 9-10; 117pp; English.
XX
CC The present sequence represents a narbonolide synthase subunit 1
CC (PICAI). The nucleotide sequence encoding it is used in the course of
CC the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 4551 AA;

Query Match 46.9%; Score 199; DB 21; Length 4551;
Best Local Similarity 48.8%; Pred. No. 3.5e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
Db 2423 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2478

Qy 61 SKATGLRLPVTLIFDHTTPAAVAARL 86
Db 2479 TRATGLQLPATLVFDHPTPLALVSLL 2504

RESULT 7
AAY67201
ID AAY67201 standard; protein; 4551 AA.
XX
AC AAY67201;
XX
DT 23-MAR-2000 (first entry)
XX
DE Narbonolide synthase subunit 1 (PICAI) protein sequence.
XX
KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 1;
KW PICAI; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS Streptomyces venezuelae.
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR N-PSDB; AAZ56001.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Example 2; Page 11-12; 98pp; English.
XX
CC This is the Streptomyces venezuelae narbonolide synthase subunit 1,
CC PICAI protein sequence. The invention relates to recombinant DNA
CC containing a coding sequence for a narbonolide polyketide synthase
CC (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC a series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC transferase enzyme is required for this conversion, and the desosamine
CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC DNA of the invention is used to express, in transformed cells,
CC narbonolide (or its derivatives) or other ketolides (particularly
CC hybrids), which may then be converted (e.g. by other enzymes
CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.
XX
SQ Sequence 4551 AA;

Query Match 46.9%; Score 199; DB 21; Length 4551;
Best Local Similarity 48.8%; Pred. No. 3.5e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
Db 2423 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2478

Qy 61 SKATGLRLPVTLIFDHTTPAAVAARL 86
Db 2479 TRATGLQLPATLVFDHPTPLALVSLL 2504

RESULT 8
AAY77192
ID AAY77192 standard; Protein; 4613 AA.
XX
AC AAY77192;

XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae macrolide biosynthetic enzyme PikAI, SEQ ID NO:31.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key Location/Qualifiers
FT Misc-difference 1386
FT /note= "Encoded by GNN"
FT Misc-difference 4581
FT /note= "Encoded by GNN"
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87297.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 19; Page 383-398; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC AAZ87295-Z87302.
XX
SQ Sequence 4613 AA;

Query Match 46.9%; Score 199; DB 21; Length 4613;
Best Local Similarity 48.8%; Pred. No. 3.6e-16;
Matches 42; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

```
Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| |||| |||      |:  |     :|  :||||: ||||||||
Db       2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2541 TRATGLQLPATLVFDHPTPLALVSLL 2566
```

RESULT 9
AAY77200
ID AAY77200 standard; Protein; 4613 AA.
XX
AC AAY77200;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster protein #1.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87318.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.

```
CC   Sequences AAY77200-Y77203 and AAY80997 represent  proteins encoded by the
CC   macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
CC   ATCC 15439, as given in figure 31.
XX
SQ   Sequence   4613 AA;

  Query Match             46.9%;  Score 199;  DB 21;  Length 4613;
  Best Local Similarity   48.8%;  Pred. No. 3.6e-16;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| ||||   |||      |:  |     :|  :||||: ||||||||
Db       2485 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 2540

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       2541 TRATGLQLPATLVFDHPTPLALVSLL 2566


RESULT 10
AAB70968
ID   AAB70968 standard; Protein; 4924 AA.
XX
AC   AAB70968;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF21, SEQ ID 48.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88338.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 57; Page 264-284; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   4924 AA;

  Query Match             46.9%;  Score 199;  DB 22;  Length 4924;
  Best Local Similarity   50.0%;  Pred. No. 3.9e-16;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | || | ::  |||::| | | : |  |     :|     |  :|| |||||:  |||||||
Db       4758 LAALPPAERGHLLLEVVLAETASTLGHD----SAEAVQPDRTFAELGFDSLTAVELRNRL 4813

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              :  |||||| ||:||| || |:: :|
Db       4814 NAVTGLRLPPTLVFDHPTPLALSEQL 4839


RESULT 11
AAY39300
ID   AAY39300 standard; Protein; 4928 AA.
XX
AC   AAY39300;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnD a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..424
FT                   /label= KS5
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                        module 5"
FT   Domain          539..866
FT                   /label= AT5
FT                   /note= "Acyl transferase domain: part of extender
FT                        module 5"
FT   Domain          893..1078
FT                   /label= DH5
FT                   /note= "Dehydratase domain: part of extender
FT                        module 5"
FT   Domain          1384..1565
FT                   /label= KR5
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                        module 5"
FT   Domain          1645..1726
FT                   /label= ACP5
FT                   /note= "Acyl carrier protein domain: part of extender
FT                        module 5"
FT   Domain          1748..2172
FT                   /label= KS6
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                        module 6"
FT   Domain          2283..2613
FT                   /label= AT6
FT                   /note= "Acyl transferase domain: part of extender
FT                        module 6"
FT   Domain          2916..3095
FT                   /label= KR6
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                        module 6"
FT   Domain          3188..3269
FT                   /label= ACP6
FT                   /note= "Acyl carrier protein domain: part of extender
FT                        module 6"
FT   Domain          3291..3713
FT                   /label= KS7
```

```
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 7"
FT   Domain          3825..4153
FT                   /label= AT7
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 7"
FT   Domain          4344..4638
FT                   /label= KR7
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 7"
FT   Domain          4725..4806
FT                   /label= ACP7
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 7"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 113-129; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnD gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnD gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   4928 AA;
```

```
  Query Match             46.9%;  Score 199;  DB 20;  Length 4928;
  Best Local Similarity   50.0%;  Pred. No. 3.9e-16;
  Matches   43;  Conservative   13;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | || | ::  |||::| | | : |  |    :|    |  :|| |||||:  |||||||
Db       4762 LAALPPAERGHLLLEVVLAETASTLGHD----SAEAVQPDRTFAELGFDSLTAVELRNRL 4817

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
                :  |||||| ||:||| || |:: :|
Db       4818 NAVTGLRLPPTLVFDHPTPLALSEQL 4843
```

```
RESULT 12
AAY77180
ID   AAY77180 standard; Protein; 12199 AA.
XX
AC   AAY77180;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87285.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 315-353; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
```

```
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents a protein
CC   encoded by the macrolide biosynthetic gene cluster (pik) from
CC   Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence   12199 AA;

  Query Match          46.9%;  Score 199;  DB 21;  Length 12199;
  Best Local Similarity   48.8%;  Pred. No. 1.4e-15;
  Matches   42;  Conservative   14;  Mismatches   26;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              | | :| ::  :|| |||   |||      |: |     :|  :|||: ||||||||
Db       3143 LAAAAPGERTEILLGLVRAQAAAVL----RMRSPEDVAADRAFKDIGFDSLAGVELRNRL 3198

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              ::||||:|| ||:||| || |: : |
Db       3199 TRATGLQLPATLVFDHPTPLALVSLL 3224


RESULT 13
AAY92707
ID   AAY92707 standard; Protein; 4150 AA.
XX
AC   AAY92707;
XX
DT   29-AUG-2000  (first entry)
XX
DE   S. antibioticus 8,8a-deoxyoleandolide synthase I encoded by oleAI.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   Region          9..1034
FT                   /label= Loading_module
FT   Domain          9..428
FT                   /label= KS-Q_domain
FT                   /note= "inactivated ketosynthase domain"
FT   Domain          562..931
FT                   /label= AT_domain
FT                   /note= "malonyl-specific acyl transferase domain"
FT   Domain          954..1034
FT                   /label= ACP_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          1061..2526
FT                   /label= Extender_module_1
FT   Domain          1061..1478
FT                   /label= KS1_domain
FT                   /note= "ketosynthase domain"
FT   Domain          1580..1926
FT                   /label= AT1_domain
FT                   /note= "acyl transferase domain"
FT   Domain          2162..2349
FT                   /label= KR1_domain
FT                   /note= "ketoreductase domain"
FT   Domain          2440..2526
FT                   /label= ACP1_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          2545..4069
FT                   /label= Extender_module_2
FT   Domain          2545..2973
FT                   /label= KS2_domain
FT                   /note= "ketosynthase domain"
FT   Domain          3075..3420
FT                   /label= AT2_domain
FT                   /note= "acyl transferase domain"
FT   Domain          3700..3884
FT                   /label= KR2_domain
FT                   /note= "ketoreductase domain"
FT   Domain          3979..4075
FT                   /label= ACP2_domain
FT                   /note= "acyl carrier protein domain"
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24478.
XX
PR   29-OCT-1998;   98US-0106100.
PR   16-FEB-1999;   99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   N-PSDB; AAA09469.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 27-28; 86pp; English.
XX
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence   4150 AA;

  Query Match          46.2%;  Score 196;  DB 21;  Length 4150;
  Best Local Similarity   50.0%;  Pred. No. 7.8e-16;
  Matches   43;  Conservative   10;  Mismatches   29;  Indels    4;  Gaps    1;

Qy          1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
              |  |:  :|: ||: |||      ||   | |  |     |  :|  |||||:  |||||||
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVA----PDRAFKDLGFDSLTSVELRNRL 2496

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              : ||||||||| :||:  |||:|  |
Db       2497 NTATGLRLPVTAVFDYARPAALAGHL 2522
```

RESULT 14
AAY67203
ID AAY67203 standard; protein; 1560 AA.
XX
AC AAY67203;
XX
DT 23-MAR-2000 (first entry)
XX
DE Narbonolide synthase subunit 3 (PICAIII) protein sequence.
XX
KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 3;
KW PICAIII; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS Streptomyces venezuelae.
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR N-PSDB; AAZ56001.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Example 2; Page 14; 98pp; English.
XX
CC This is the Streptomyces venezuelae narbonolide synthase subunit 3,
CC PICAIII protein sequence. The invention relates to recombinant DNA
CC containing a coding sequence for a narbonolide polyketide synthase
CC (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC a series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC transferase enzyme is required for this conversion, and the desosamine
CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC DNA of the invention is used to express, in transformed cells,
CC narbonolide (or its derivatives) or other ketolides (particularly
CC hybrids), which may then be converted (e.g. by other enzymes
CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.
XX
SQ Sequence 1560 AA;

Query Match 46.0%; Score 195; DB 21; Length 1560;
Best Local Similarity 48.8%; Pred. No. 2.8e-16;
Matches 42; Conservative 11; Mismatches 29; Indels 4; Gaps 1;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
:||| ::: || ||| | ||| :| | | :| ||||| |:|||:|
Db 1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy 61 SKATGLRLPVTLIFDHTTPAAVAARL 86
| | ||| | :||| |||||:|| |
Db 1449 STVVGNRLPATTVFDHPTPAALAAHL 1474

RESULT 15
AAB18639
ID AAB18639 standard; Protein; 1562 AA.
XX
AC AAB18639;
XX
DT 22-JAN-2001 (first entry)
XX
DE Amino acid sequence of narbonolide synthase subunit 3 (PICAIII).
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 11-14; 117pp; English.
XX
CC The present sequence represents a narbonolide synthase subunit 3
CC (PICAIII). The nucleotide sequence encoding it is used in the course of
CC the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 1562 AA;

Query Match 46.0%; Score 195; DB 21; Length 1562;
Best Local Similarity 48.8%; Pred. No. 2.8e-16;
Matches 42; Conservative 11; Mismatches 29; Indels 4; Gaps 1;

Qy 1 LTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 60
:||| ::: || ||| | ||| :| | | :| ||||| |:|||:|

```
Db       1393 ITALPEPERRPALLTLVRTHAAAVLGHSSPDRVA----PGRAFTELGFDSLTAVQLRNQL 1448

Qy         61 SKATGLRLPVTLIFDHTTPAAVAARL 86
              |   | ||| | :||| ||||:|| |
Db       1449 STVVGNRLPATTVFDHPTPAALAAHL 1474
```

Search completed: June 17, 2003, 13:02:28
Job time : 5.21291 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 8.22676 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_3537_3714
Perfect score: 882
Sequence: 1 GSVLVTGGTGVLGAAVARHL.....LAYRRRAAGLPGVSLAWGLW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	882	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	882	100.0	5532	2	Q9S0R4	Q9s0r4 streptomyce
3	616	69.8	6239	2	Q9S0R7	Q9s0r7 streptomyce
4	564	63.9	2034	2	Q93NX7	Q93nx7 streptomyce
5	561	63.6	4151	16	O53490	O53490 mycobacteri
6	554	62.8	2066	2	Q9L4X1	Q9l4x1 streptomyce
7	554	62.8	3654	2	O30766	O30766 streptomyce
8	549	62.2	11096	2	Q9L4W3	Q9l4w3 streptomyce
9	548.5	62.2	9477	2	Q9L4X3	Q9l4x3 streptomyce
10	547	62.0	3739	2	Q9ZGI4	Q9zgi4 streptomyce
11	546	61.9	4613	2	Q9ZGI5	Q9zgi5 streptomyce
12	546	61.9	5435	2	Q9L4X2	Q9l4x2 streptomyce
13	544	61.7	3613	2	Q93HJ1	Q93hj1 streptomyce
14	538	61.0	10917	2	Q93NW6	Q93nw6 streptomyce
15	537	60.9	902	2	Q93AL5	Q93al5 streptomyce
16	535	60.7	6315	2	Q9ADL6	Q9adl6 polyangium
17	535	60.7	9507	2	Q9EWA1	Q9ewa1 streptomyce
18	534	60.5	4472	2	O33954	O33954 streptomyce
19	532	60.3	5644	2	Q93NX8	Q93nx8 streptomyce
20	531	60.2	4498	2	Q93HE5	Q93he5 streptomyce
21	526	59.6	1808	2	Q9EWA2	Q9ewa2 streptomyce
22	522	59.2	1907	2	Q93HI9	Q93hi9 streptomyce
23	521	59.1	3576	16	Q9EX53	Q9ex53 streptomyce
24	520	59.0	3939	2	Q93HJ3	Q93hj3 streptomyce
25	520	59.0	4735	2	O54666	O54666 amycolatops
26	520	59.0	5588	2	Q9ALM2	Q9alm2 saccharopol
27	519	58.8	8817	2	Q53840	Q53840 polyangium
28	519	58.8	9510	2	Q93NX9	Q93nx9 streptomyce
29	518	58.7	1864	2	O33955	O33955 streptomyce
30	512	58.0	1835	2	Q93H85	Q93h85 streptomyce
31	512	58.0	3524	2	Q93H86	Q93h86 streptomyce
32	511	57.9	6797	2	Q9X993	Q9x993 streptomyce
33	509	57.7	2024	2	Q9EWA3	Q9ewa3 streptomyce
34	508	57.6	2152	16	Q9EX54	Q9ex54 streptomyce
35	504.5	57.2	1839	2	O30765	O30765 streptomyce
36	504	57.1	4928	2	Q9ALM3	Q9alm3 saccharopol
37	503.5	57.1	3413	2	O54593	O54593 amycolatops
38	500.5	56.7	6048	2	Q93H87	Q93h87 streptomyce
39	496	56.2	4340	2	O30764	O30764 streptomyce
40	495.5	56.2	1728	2	O54591	O54591 amycolatops
41	493.5	56.0	6146	2	Q93HJ5	Q93hj5 streptomyce
42	492	55.8	5060	2	O52545	O52545 amycolatops
43	492	55.8	5069	2	O52789	O52789 amycolatops
44	491	55.7	2731	2	Q9Z3T9	Q9z3t9 pseudomonas
45	491	55.7	4881	2	Q9S0R3	Q9s0r3 streptomyce

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 882; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1.1e-54;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3537 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 3596

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3597 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 3656

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3657 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 3714


RESULT 2
Q9S0R4
ID   Q9S0R4      PRELIMINARY;      PRT;  5532 AA.
AC   Q9S0R4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 3.
GN   AVEA3.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84478.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5532 AA;  575193 MW;  EC7B2069D44A4057 CRC64;

  Query Match             100.0%;  Score 882;  DB 2;  Length 5532;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-54;
  Matches  178;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1488 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 1547

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1548 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 1607

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1608 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 1665


RESULT 3
Q9S0R7
ID   Q9S0R7      PRELIMINARY;      PRT;  6239 AA.
AC   Q9S0R7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 2.
GN   AVEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84475.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6239 AA;  666283 MW;  A706280940B502FA CRC64;

  Query Match             69.8%;  Score 616;  DB 2;  Length 6239;
  Best Local Similarity   69.1%;  Pred. No. 2.3e-35;
  Matches  123;  Conservative   16;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||||||  |||||    ||| |||  ||||||  ||  |||||  ||| | :||
Db       5753 GTVLITGGTGVLGRLVARHLVEAHGVRHLLLAGRRGPDAEGAPELRAELGGLGATVEVVA 5812

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|:::  ||  :|   |||||||:||:|||  | ||:|||||  |  || |||||||
Db       5813 CDAADRQQLADLLTRIPDDRPLTGVVHSAGILDDGVITSLSPERLGAVLRAKADAALLLD 5872

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||||  ||||||:||||: ::|| ||||||||||  || ||:|||| ||| ||||||||
Db       5873 ELTRGAELSAFVMFSSASAVVGSPGQGNYAAANAVLDFLAHRRRAEGLPAVSLAWGLW 5930


RESULT 4
Q93NX7
ID   Q93NX7      PRELIMINARY;      PRT;  2034 AA.
AC   Q93NX7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphK.
GN   AMPHK.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73503.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
```

```
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   2034 AA;  212063 MW;  83E962B7DC3D5747 CRC64;

  Query Match           63.9%;  Score 564;  DB 2;  Length 2034;
  Best Local Similarity   65.2%;  Pred. No. 3.6e-32;
  Matches  116;  Conservative   17;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||||:||||    | | ||| ||||||||   | |||  |||:| : |
Db       1386 GTVLITGGTGGLGAALARHLVTSRGARHLLLASRRGPDAPGTSELVAELTGLGAQVTVSA 1445

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||:|  |  |:  |||  ||| ||| |||||||   |||||:|  |   | |||| |
Db       1446 CDVGDRDAVDTLVASVPAEHPLTAVVHTAGVLDDALTGSLTPEQLAGVLRPKADAALHLH 1505

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              | | | :|:||||:|| :|::|  || |||||||:|||||:||||||||:||||| |
Db       1506 EATLGQDLAAFVLYSSISGVIGGPGQANYAAANASLDALAHRRRAAGLPGLSLAWGPW 1563


RESULT 5
O53490
ID   O53490      PRELIMINARY;      PRT;  4151 AA.
AC   O53490;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase.
GN   PKS12 OR RV2048C OR MT2108 OR MTV018.35C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021899; CAA17262.1; -.
DR   EMBL; AE007061; AAK46387.1; -.
DR   HSSP; P25715; 1MLA.
DR   TIGR; MT2108; -.
DR   TuberculList; Rv2048c; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine; Transferase; Complete proteome.
FT   CONFLICT   2147   2147       H -> Q (IN REF. 2).
FT   CONFLICT   2260   2261       DI -> TV (IN REF. 2).
FT   CONFLICT   2268   2268       W -> H (IN REF. 2).
FT   CONFLICT   2272   2272       S -> A (IN REF. 2).
FT   CONFLICT   2279   2280       AF -> PY (IN REF. 2).
FT   CONFLICT   2282   2283       AA -> GR (IN REF. 2).
FT   CONFLICT   2287   2287       T -> V (IN REF. 2).
FT   CONFLICT   2289   2289       F -> W (IN REF. 2).
FT   CONFLICT   3004   3004       S -> L (IN REF. 2).
FT   CONFLICT   3649   3649       P -> A (IN REF. 2).
SQ   SEQUENCE   4151 AA;  431577 MW;  536F644F111D7A5D CRC64;

  Query Match           63.6%;  Score 561;  DB 16;  Length 4151;
  Best Local Similarity   60.7%;  Pred. No. 1.3e-31;
  Matches  108;  Conservative   31;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||: |:|||||:    |||:|:||||||||||  | ||||| ||:|::||
Db       1680 GTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVAELAAAGAQVQVVA 1739

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|  : :::  :|   ||:||:| || |||| : ||||:|:  |  :|||||  |
Db       1740 CDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRVDVVLRSKVDAAWHLH 1799

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| :::||||:||| ||::||:|| ||||||: |||||  ||| ||| :|| ||||
Db       1800 ELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALAAHRRAHGLPAISLGWGLW 1857


RESULT 6
Q9L4X1
ID   Q9L4X1      PRELIMINARY;      PRT;  2066 AA.
AC   Q9L4X1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysK.
GN   NYSK.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71768.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
```

```
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2066 AA;  214572 MW;  CB02F5991FB2B56E CRC64;

  Query Match             62.8%;  Score 554;  DB 2;  Length 2066;
  Best Local Similarity   64.0%;  Pred. No. 1.9e-31;
  Matches  114;  Conservative   18;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||||:||||  | | | |||  ||||:||||  | ||| | ||:||: |
Db       1416 GTVLITGGTGGLGAALARHLVTVRGARHLLLAGRRGPEAPGAGELVAELTAQGADVRVAA 1475

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||:|  :  ||  |||  ||| ||| ||||||| | ||||::| ||   | |||  |
Db       1476 CDVGDRTALDALLATVPAAHPLTAVVHTAGVLDDALIGSLTPDQLATVLRPKADAAWHLH 1535

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              : |||::|: |||:|| :|:||| |||||||||| |||||   |   ||| :||||| |
Db       1536 DATRGLDLAGFVLYSSVSGVLGSPGQGNYAAANAYLDALARHRADQGLPALSLAWGPW 1593


RESULT 7
O30766
ID   O30766      PRELIMINARY;      PRT;  3654 AA.
AC   O30766;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase modules 4 and 5.
GN   NIDA3.
OS   Streptomyces caelestis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=36816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL-2821;
RX   MEDLINE=98053867; PubMed=9393718;
RA   Kakavas S.J., Katz L., Stassi D.;
RT   "Identification and characterization of the niddamycin polyketide
RT   synthase genes from Streptomyces caelestis.";
RL   J. Bacteriol. 179:7515-7522(1997).
DR   EMBL; AF016585; AAC46026.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001827; Antennapedia.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00032; ANTENNAPEDIA; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3654 AA;  379108 MW;  59825E49EDCD06FF CRC64;

  Query Match             62.8%;  Score 554;  DB 2;  Length 3654;
  Best Local Similarity   64.6%;  Pred. No. 3.6e-31;
  Matches  115;  Conservative   17;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  :|: :||||     ||| ||| || |||  || || ||||  ||   :||
Db       3192 GTVLITGGTGTIGSRIARHLVTRHGVRHLLLASRGGPDGDGAAGLVAELADAGASATVVA 3251

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :   | |||: ||   ||| |||:||||||  : |||||||. |   |||||: ||
Db       3252 CDVADADAVRRLLDEVPEQQPLTAVVHSAGVLDDGMLPSLTPERLERVLRPKVDAAVHLD 3311

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||| ::|||||||||:| :|||  |||||||| ||||| |||: ||| ||:|||||
Db       3312 RLTRDLDLSAFVLFSSSAALLGSPAQGNYAAANAVLDALAARRRSLGLPAVSVAWGLW 3369


RESULT 8
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match             62.2%;  Score 549;  DB 2;  Length 11096;
  Best Local Similarity   65.2%;  Pred. No. 2.9e-30;
  Matches  116;  Conservative   12;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  |||  ||||     |||:|||||||||||||    | ||| | |||| : |
Db       8812 GTVLITGGTGGLGAQFARHLVDRYGVRNLLLVSRRGPDAPGTTELVAELTAHGAEVAVQA 8871

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :   |  |: |||   ||  ||| ||||||  | ||| ||| ||   | |||  |
Db       8872 CDVADGDAVAALVAGVPDEHPLRAVVHTAGVLDDGVIGSLTEERLATVLRPKADAAWHLH 8931

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              | |||::| |||:||| ||: | ||| |||||||  |||| :|||||||||:||||| |
```

```
Db       8932 EATRGLDLDAFVVFSSVAGVFGGAGQANYAAANAFLDALMAQRRAAGLPGLSLAWGPW 8989


RESULT 9
Q9L4X3
ID   Q9L4X3      PRELIMINARY;      PRT;  9477 AA.
AC   Q9L4X3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysI.
GN   NYSI.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71766.1; -.
DR   HSSP; P81989; 1QE5.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   9477 AA;  988132 MW;  0D2BCA5D7B265483 CRC64;

  Query Match             62.2%;  Score 548.5;  DB 2;  Length 9477;
  Best Local Similarity   63.5%;  Pred. No. 2.6e-30;
  Matches  113;  Conservative   21;  Mismatches   43;  Indels    1;  Gaps    1;

Qy         1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
             |:||||||| ||| ||| ||   | : |:|:||||||||||  ||||| ||| :| : |
Db      9050 GTVLVTGGTGALGAHVARRLA-KDGAQHLVLLSRRGPDAPGAAELRAELDALGTDVTVAA 9108

Qy        61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
             ||| :| ::  :|: :||  ||||||| ||||||  :  |||||   || ||| :|||||
Db      9109 CDVADRDQLTAVLDALPADRPLTGVVHTAGVLDDGVLDRLTPERFQEVFRAKVTSALLLD 9168

Qy       121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
             ||||  ||:|||||||:  :|: || ||||||| ||||| :||  |||  |::|| |
Db      9169 ELTRDRELAAFVLFSSASAAVGNPGQANYAAANAVLDALAEQRRVLGLPATSVSWGAW 9226


RESULT 10
Q9ZGI4
ID   Q9ZGI4      PRELIMINARY;      PRT;  3739 AA.
AC   Q9ZGI4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAII.
GN   PIKAII.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR   EMBL; AF079138; AAC69330.1; -.
DR   HSSP; Q02054; 2AF8.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3739 AA;  387170 MW;  3D3910824DA5B080 CRC64;

  Query Match             62.0%;  Score 547;  DB 2;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 1.2e-30;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy         1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
             |:||:||||| ||  |||| |  ||| |||||||| |||||  |  || ||||:| : |
Db      3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy        61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
             ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db      3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy       121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
             |||   | :|:|||:||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db      3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 11
Q9ZGI5
ID   Q9ZGI5      PRELIMINARY;      PRT;  4613 AA.
AC   Q9ZGI5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAI.
GN   PIKAI.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
```

```
DR   EMBL; AF079138; AAC69329.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4613 AA;  477281 MW;  A3BAF8D37CEC9383 CRC64;

  Query Match             61.9%;  Score 546;  DB 2;  Length 4613;
  Best Local Similarity   62.8%;  Pred. No. 1.8e-30;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||: |  ||| |||||||| |||||: |  || ||||:| : |
Db       4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||     :|:|||:||||| : | |||| |||||| |||||:|||||||| :|| ||||
Db       4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279


RESULT 12
Q9L4X2
ID   Q9L4X2      PRELIMINARY;      PRT;  5435 AA.
AC   Q9L4X2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysJ.
GN   NYSJ.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71767.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5435 AA;  562659 MW;  AA55465DF087A38C CRC64;

  Query Match             61.9%;  Score 546;  DB 2;  Length 5435;
  Best Local Similarity   62.4%;  Pred. No. 2.1e-30;
  Matches  111;  Conservative   17;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  :||||     |:: |||  |||||||||  || |||||||| : |
Db       4997 GTVLITGGTGGLGGLLARHLVTGHGIKHLLLAGRRGPDAPGARALRDELAALGAEVTVAA 5056

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  : |||  :|    ||| ||| |||||||| :|||||| ||  || |||  |
Db       5057 CDVADRAALDRLLAQLPPEHPLTAVVHTAGVLDDATVGTLTPERLDTVLRAKADAAWHLH 5116

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              : ||  :|: |||:|| ||: |  |||||| |  |||||  | | ||||:||||| |
Db       5117 DATRDRDLAGFVLYSSVAGVTGGPGQGNYAAGNTFLDALAAHRAAQGLPGLSLAWGPW 5174


RESULT 13
Q93HJ1
ID   Q93HJ1      PRELIMINARY;      PRT;  3613 AA.
AC   Q93HJ1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA7.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69196.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3613 AA;  379236 MW;  6FAAE5CAB158E442 CRC64;

  Query Match             61.7%;  Score 544;  DB 2;  Length 3613;
  Best Local Similarity   64.0%;  Pred. No. 1.9e-30;
  Matches  114;  Conservative   13;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:|||||||| ||| |||||     ||| |:|  |||  ||||  | |||  |||: ||:|
Db       2940 GTVLVTGGTGALGAVVARHLVTAHGVRHLVLAGRRGERAPGAAALLAELTELGADTRILA 2999
```

```
Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||| :|  :  ||  :||  ||| ||| |||| | |: ||||:|  ||  || |||  |
Db        3000 CDVSDRDALAVLLRDIPADRPLTAVVHTAGVLADGTVESLTPDRFDTVLRAKADAAWHLH 3059

Qy         121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               |||:   |  |||||||||:||| |||||||||: |||||  || ||||| :|||| |
Db        3060 ELTQDAPLREFVLFSSAAGLLGSQGQGNYAAANSFLDALAAHRREAGLPGTALAWGWW 3117


RESULT 14
Q93NW6
ID   Q93NW6      PRELIMINARY;      PRT; 10917 AA.
AC   Q93NW6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphC.
GN   AMPHC.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73514.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00059; ADH_ZINC; UNKNOWN_1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
KW   Phosphopantetheine.
SQ   SEQUENCE   10917 AA;  1132905 MW;  15AC5956BB5810A1 CRC64;

  Query Match             61.0%;  Score 538;  DB 2;  Length 10917;
  Best Local Similarity   63.5%;  Pred. No. 1.7e-29;
  Matches  113;  Conservative   13;  Mismatches   52;  Indels    0;  Gaps    0;

Qy           1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
               |:|||||||| |||  ||||    |||:|||||||| ||||   | ||| | |||| : |
Db        8695 GTVLVTGGTGGLGAQFARHLVDAYGVRNLLLVSRRGGDAPGTTELVAELIAHGAEVTVQA 8754

Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||| :   |  |: |:||  ||| ||| ||||||  | ||| |||  |   | |||  |
Db        8755 CDVADPDAVTALVAGIPAEHPLTAVVHTAGVLDDGVIGSLTGERLARVLRPKADAAWNLH 8814

Qy         121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               : |||::| ||::||| ||: | ||| |||| || ||||   |||||||:|||| |
Db        8815 QATRGLDLDAFIVFSSVAGVFGGAGQANYAAGNAFLDALMEHRRAAGLPGLSLAWGPW 8872


RESULT 15
Q93AL5
ID   Q93AL5      PRELIMINARY;      PRT;   902 AA.
AC   Q93AL5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Rearranged polyketide synthase (Fragment).
GN   NYSCD48.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ERD48;
RA   Brautaset T., Bruheim P., Sletta H., Hagen L., Ellingsen T.E.,
RA   Strom A.R., Valla S., Zotchev S.B.;
RT   "Hexaene derivative of the polyene antifungal antibiotic nystatin is
RT   produced as a result of an induced rearrangement within the nysC
RT   polyketide synthase gene in Streptomyces noursei ATCC 11455.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF420594; AAL18806.1; -.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PRINTS; PR01610; CD36ANTIGEN.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT   NON_TER       1      1
FT   NON_TER     902    902
SQ   SEQUENCE   902 AA;  93381 MW;  C1888F648C12C7CC CRC64;

  Query Match             60.9%;  Score 537;  DB 2;  Length 902;
  Best Local Similarity   62.9%;  Pred. No. 1.3e-30;
  Matches  112;  Conservative   11;  Mismatches   55;  Indels    0;  Gaps    0;

Qy           1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
               |:|||||||| ||  |||||    |||| ||| || |  | || || ||| :||| | : |
Db         391 GTVLVTGGTGGLGGLVARHLVRSCGVRHLLLTSRSGVGAAGAAGLVAELESLGARVVVAA 450

Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||||:   |  |: ||    ||: |||||||||| : |||||||  |   ||| |  |
Db         451 CDVGDGSAVAELVAGVSESYPLSAVVHAAGVLDDGVVGSLTPERLAAVLRPKVDGAWNLH 510

Qy         121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               | |||::| |||:||| ||: | ||| |||| || ||||   | | |||||||||| |
Db         511 EATRGLDLDAFVVFSSVAGVFGGAGQANYAAGNAFLDALMVHRVAGGLPGVSLAWGAW 568


Search completed: June 17, 2003, 13:07:58
Job time : 9.22676 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 1.94115 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_3537_3714
Perfect score: 882
Sequence: 1 GSVLVTGGTGVLGAAVARHL.....LAYRRRAAGLPGVSLAWGLW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	475	53.9	3567	1	ERY2_SACER	Q03132 saccharopol
2	427	48.4	3172	1	ERY3_SACER	Q03133 saccharopol
3	426.5	48.4	3491	1	ERY1_SACER	Q03131 saccharopol
4	392	44.4	3519	1	OL56_STRAT	Q07017 streptomyce
5	295.5	33.5	2110	1	MCAS_MYCBO	Q02251 mycobacteri
6	249	28.2	1774	1	MSAS_PENPA	P22367 penicillium
7	248	28.1	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	248	28.1	4447	1	PKSK_BACSU	P40803 bacillus su
9	237	26.9	4427	1	PKSL_BACSU	Q05470 bacillus su
10	225.5	25.6	1538	1	PPSB_MYCTU	Q10978 mycobacteri
11	225.5	25.6	2511	1	FAS_CHICK	P12276 gallus gall
12	214.5	24.3	2504	1	FAS_HUMAN	P49327 homo sapien
13	207.5	23.5	2505	1	FAS_RAT	P12785 rattus norv
14	186.5	21.1	838	1	FAS_MOUSE	P19096 mus musculu
15	169	19.2	4273	1	PKSM_BACSU	P40872 bacillus su
16	150.5	17.1	255	1	FABG_MYCAV	O07399 mycobacteri
17	140.5	15.9	255	1	FABG_MYCSM	P71534 mycobacteri
18	135.5	15.4	247	1	FABG_MYCTU	Q48930 mycobacteri
19	134	15.2	251	1	YMP3_STRCO	P43168 streptomyce
20	133.5	15.1	245	1	NODG_RHIS3	P72332 rhizobium s
21	133.5	15.1	247	1	YD50_MYCTU	Q11020 mycobacteri
22	132	15.0	247	1	FABG_PSEAE	O54438 pseudomonas
23	129	14.6	247	1	YGFF_ECOLI	P52037 escherichia
24	128	14.5	242	1	FABG_ACTAC	P70720 actinobacil
25	125.5	14.2	245	1	NODG_RHIME	P06234 rhizobium m
26	124.5	14.1	320	1	FABG_CUPLA	P28643 cuphea lanc
27	115	13.0	246	1	PHBB_CHRVI	P45375 chromatium
28	112	12.7	261	1	DHB8_HUMAN	Q92506 homo sapien
29	111.5	12.6	261	1	SPRE_MOUSE	Q64105 mus musculu
30	110.5	12.5	247	1	FAG1_SYNY3	P73574 synechocyst
31	110	12.5	241	1	PHBB_RHIME	P50205 rhizobium m
32	110	12.5	241	1	PHBB_ZOORA	P23238 zoogloea ra
33	110	12.5	242	1	PHAB_PARDE	P50204 paracoccus
34	109.5	12.4	246	1	NODG_AZOBR	P17611 azospirillu
35	109.5	12.4	260	1	DHB8_MOUSE	P50171 mus musculu
36	109	12.4	736	1	DHB4_HUMAN	P51659 homo sapien
37	105	11.9	248	1	PHAB_ACISP	P50203 acinetobact
38	104.5	11.8	336	1	TS2_MAIZE	P50160 zea mays (m
39	103.5	11.7	262	1	SPRE_RAT	P18297 rattus norv
40	102.5	11.6	244	1	FABG_ECOLI	P25716 escherichia
41	101.5	11.5	247	1	FABG_CHLTR	P38004 chlamydia t
42	101.5	11.5	250	1	LINC_PSEPA	P50197 pseudomonas
43	101	11.5	240	1	FAG2_SYNY3	P73826 synechocyst
44	101	11.5	254	1	Y1J1_MYCTU	P72057 mycobacteri
45	101	11.5	735	1	DHB4_RAT	P97852 rattus norv

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             53.9%;  Score 475;  DB 1;  Length 3567;
  Best Local Similarity   57.9%;  Pred. No. 3.1e-28;
  Matches  103;  Conservative   17;  Mismatches   56;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  :||||    ||| |||||||| ||||:: ||||:  |||   | |
Db       3141 GTVLITGGTGTLGRLLARHLVTEHGVRHLLLVSRRGADAPGSDELRAEIEDLGASAEIAA 3200

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|  :  ||:|:|   |||||||||||| |  : |:    :  |  ||||||  |
Db       3201 CDTADRDALSALLDGLPR--PLTGVVHAAGVLADGLVTSIDEPAVEQVLRAKVDAAWNLH 3258

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||    || |||||||| :|   ||| ||||| :|:|||  ||  |||  :| ||||
Db       3259 ELTANTGLSFFVLFSSAASVLAGPGQGVYAAANESLNALAALRRTRGLPAKALGWGLW 3316


RESULT 2
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
```

```
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match           48.4%;  Score 427;  DB 1;  Length 3172;
  Best Local Similarity   55.9%;  Pred. No. 1.2e-24;
  Matches  100;  Conservative   17;  Mismatches   56;  Indels    6;  Gaps    4;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |: ||||||| |||  ||||||  ||| ||:|||||| ||| |  | ||| ||| :  | |

Db       2557 GTALVTGGTGALGGHVARHLAR-CGVEDLVLVSRRGVDAPAAAELEAELVALGPKTTITA 2615

Qy         61 CDVGERREVVRLLEGVPA-GCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
              ||| :| :: :||| :   | |:  ||| ||| :   :  :    | :| ||||  | ||
Db       2616 CDVADREQLSKLLEELRGQGRPVRTVVHTAGVPESRPLHEI--GELESVCAAKVTGARLL 2673

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||    |   |||||| ||: |||  | |:|||| |||||:|||| |    |:||| |
Db       2674 DELCPDAE--TFVLFSSGAGVWGSANLGAYSAANAYLDALAHRRRAEGRAATSVAWGAW 2730


RESULT 3
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
```

```
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match           48.4%;  Score 426.5;  DB 1;  Length 3491;
  Best Local Similarity   53.9%;  Pred. No. 1.4e-24;
  Matches   96;  Conservative   18;  Mismatches   63;  Indels    1;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| :|  :|| ||   |   |||||| |||| ||  | ||| ||||   : |
Db       1613 GTVLVTGGTGGVGGQIARWLAR-RGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAA 1671

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  |  || |:    ||: | ||| ||| |: :|| ||:     |||  |  |
Db       1672 CDVTDRESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLH 1731

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::|:|||||| |   |: | | ||  || || || :||: |||  ::||| |
Db       1732 ELTRELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTW 1789


RESULT 4
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match           44.4%;  Score 392;  DB 1;  Length 3519;
  Best Local Similarity   53.1%;  Pred. No. 5.8e-22;
  Matches   95;  Conservative   16;  Mismatches   64;  Indels    4;  Gaps    3;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAAL-GAEVRIV 59
              |:|||||| | |||  || | |  |   ::|||||  ||||  |  ||  | || | :
Db       1202 GTVLVTGGLGGLGAHTARWLVG-GGADHVVLVSRRGGSAPGAGDLVRELEGLGGARVSVR 1260

Qy         60 ACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
              |||| :|  :  ||  :  | |:| | |||||     :| :: :    | ||||  |: |
Db       1261 ACDVADRVALRALLSDL--GEPVTAVFHAAGVPQSTPLAEISVQEAADVMAAKVAGAVNL 1318

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               ||      | ||||||| ||: || ||  |||||| ||||| |||  |||  |:|||:|
Db       1319 GELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPATSVAWGMW 1377


RESULT 5
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
```

```
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             33.5%;  Score 295.5;  DB 1;  Length 2110;
  Best Local Similarity   41.3%;  Pred. No. 7e-15;
  Matches   76;  Conservative   29;  Mismatches   66;  Indels   13;  Gaps    6;

Qy          1 GSVLVTGGTGVLGAAVARHLAGV-CGVRDLLLVSRRGPDAPGA----EGLRAELAALGAE 55
              || ::||| | ||   |  ||   ||   ::| :|  |: | |    ||||    | ||:
Db       1764 GSYIITGGLGGLGLFFASKLAAAGCG--RIVLTARSQPN-PKARQTIEGLR----AAGAD 1816

Qy         56 VRIVACDVGERREVVRLLEGVPA-GCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVD 114
                :  ::   :: |     ||:    | | || ||:|:| |::|||: ::| | :   :: ||
Db       1817 IVVECGNIAEPDTADRLVSAATATGLPLRGVLHSAAVVEDATLTNITDELIDRDWSPKVF 1876

Qy        115 AALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLA 174
               :  |   | |  |  | |||| | :||| ||| ||||| : :|  |:  || |||  ::|
Db       1877 GSWNLHRATLGQPLDWFCLFSSGAALLGSPGQGAYAAANSWVDVFAHWPRAQGLPVSAIA 1936

Qy        175 WGLW 178
               || |
Db       1937 WGAW 1940


RESULT 6
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
```

```
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             28.2%;  Score 249;  DB 1;  Length 1774;
  Best Local Similarity   35.7%;  Pred. No. 1.9e-11;
  Matches   70;  Conservative   26;  Mismatches   76;  Indels   24;  Gaps    6;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPG-------AEGLRAELAAL- 52
              |: |:||| |||| || |  | | |||:|||    |          :| |:  :| :
Db       1412 GTYLITGGLGVLGLEVADFLVEK-GARRLLLISRRA--LPPRRTWDQVSEDLQPTIAKIR 1468

Qy         53 -----GAEVRIVACDVGERREVVRLLE-----GVPAGCPLTGVVHAAGVLDDATIASLTP.102
                  || | ::  |: :   | :|        :|:   : ||||||||||:  :   |
Db       1469 LLESRGASVHVLPLDITKPDAVEQLTTALDRLSLPS---VQGVVHAAGVLDNELVMQTTR 1525

Qy        103 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 162
              :   | | |:  || | |:     :  ||:|||   ::|  || :| : || || ||
Db       1526 DAFNRVLAPKIAGALALHEVFPPKSVDFFVMFSSCGNLVGFTGQASYGSGNAFLDTLATH 1585

Qy        163 RRAAGLPGVSLAWGLW 178
              |  |   ||  |  |
Db       1586 RARLGDAAVSFQWTSW 1601


RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             28.1%;  Score 248;  DB 1;  Length 1876;
  Best Local Similarity   36.1%;  Pred. No. 2.4e-11;
  Matches   69;  Conservative   27;  Mismatches   77;  Indels   18;  Gaps    4;

Qy          4 LVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAE--------GLRAELAAL--- 52
              |:||| | ||  :|  ||      |  |::: | |  |  :          ||  : |:
Db       1494 LITGGLGALGLLMADWLADRGAHR--LVLTGRTPLPPRRDWQLDTLDTELRRRIDAIRAL 1551

Qy         53 ---GAEVRIVACDVGERREVVRLL--EGVPAGCPLTGVVHAAGVLDDATIASLTPERLGT 107
                 |  |  || ||| | :|  ||         |: |::||||: :|  : |:| : :
Db       1552 EMRGVTVEAVAADVGCREDVQALLAARDRDGAAPIRGIIHAAGITNDQLVTSMTGDAVRQ 1611

Qy        108 VFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAG 167
              |   |:  : :| :      :  | | :|||| |  |||:||||: ||||  || |
Db       1612 VMWPKIGGSQVLHDAFPPGSVDFFYLTASAAGIFGIPGQGSYAAANSYLDALARARRQQG 1671

Qy        168 LPGVSLAWGLW 178
               :|| |  |
Db       1672 CHTMSLDWVAW 1682


RESULT 8
PKSK_BACSU
ID   PKSK_BACSU     STANDARD;      PRT;  4447 AA.
AC   P40803;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksK (PKS).
GN   PKSK.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=95219083; PubMed=7704258;
RA   Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;
RT   "Sequence around the 159 degree region of the Bacillus subtilis
RT   genome: the pksX locus spans 33.6 kb.";
```

```
RL   Microbiology 141:299-309(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11039; AAA85144.1; -.
DR   EMBL; Z99112; CAB13590.1; -.
DR   EMBL; Z99113; CAB13601.1; -.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10930; pksK.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF00668; Condensation; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;
KW   Complete proteome.
FT   DOMAIN       1     68       ACYL CARRIER (ACP) 1.
FT   DOMAIN    1063   1130       ACYL CARRIER (ACP) 2.
FT   DOMAIN    2516   2589       ACYL CARRIER (ACP) 3.
FT   DOMAIN    2618   2687       ACYL CARRIER (ACP) 4.
FT   DOMAIN    3868   3937       ACYL CARRIER (ACP) 5.
FT   BINDING     31     31       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   1093   1093       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   2552   2552       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   2650   2650       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE  2915   2915       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING   3900   3900       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE  4147   4147       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
SQ   SEQUENCE   4447 AA;  496058 MW;  9C6DB4A6C0A9C057 CRC64;

  Query Match             28.1%;  Score 248;  DB 1;  Length 4447;
  Best Local Similarity   36.2%;  Pred. No. 5.6e-11;
  Matches   67;  Conservative   25;  Mismatches   85;  Indels    8;  Gaps    4;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |  |:||| | ||   |: :|   | |  :::: |   :   |     | ::||||
Db       2254 GVYLITGGAGSLGLLFAKEIANRTG-RSTIVLTGRSVLSEDKENELEALRSIGAEVVYRE 2312

Qy         61 CDVGERREVVRLLEGVPAGC-PLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
              || ::  |  ||| :      | |::| ||   |  |  | |    |   ||   | :
Db       2313 ADVSDQHAVRHLLEEIKERYGTLNGIIHGAGSSKDRFIIHKTNEEFQEVLQPKVSGLLHV 2372

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALA-YRRRAAGL-----PGVSL 173
              || ::   |  |: ||| :| ||:||| :|||||: :|| | |||  |        :|
Db       2373 DECSKDFPLDFFIFFSSVSGCLGNAGQADYAAANSFMDAFAEYRRSLAASKKRFGSTISF 2432

Qy        174 AWGLW 178
               | ||
Db       2433 NWPLW 2437


RESULT 9
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
```

```
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;

  Query Match             26.9%;  Score 237;  DB 1;  Length 4427;
  Best Local Similarity   36.2%;  Pred. No. 3.8e-10;
  Matches   67;  Conservative   28;  Mismatches   82;  Indels    8;  Gaps    4;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |  |:||| | ||   | :|       ::|  |   |    : |:| |  || :
Db       1410 GVYLITGGAGGLGFIFATEIANQTNDAVVILTGRSPLDERKKKKLKA-LQKLGIQAIYRQ 1468

Qy         61 CDVGERREVVRLL-EGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
               |: ::: |  || |       | |::|:||:: |  |     | : || | ||   : |
Db       1469 ADLADKQTVDALLKETQNVYGDLDGIIHSAGLIKDNFIMKKKKEEVQTVLAPKVAGLIHL 1528

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALA-YRRRAAGL-----PGVSL 173
              || |: : |  |:|||| || :||||| :|| ||| ::| : ||   | |      :|:
Db       1529 DEATKDIPLDFFILFSSGAGAVGSAGQADYAMANAFMNAFSEYRNGQAELHKRYGKTLSV 1588

Qy        174 AWGLW 178
              | ||
Db       1589 CWPLW 1593


RESULT 10
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             25.6%;  Score 225.5;  DB 1;  Length 1538;
  Best Local Similarity   35.2%;  Pred. No. 1e-09;
  Matches   64;  Conservative   32;  Mismatches   73;  Indels   13;  Gaps    5;

Qy          2 SVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA-EGLRAELAALGAEVRIVA 60
              | || | || :|  : | || : | : :: ::|:    ||| : |   ||| | ::  ||
Db       1153 SQLVIGATGNIGPHLIRQLARM-GAKTIVAMARK----PGALDELTQCLAATGTDLIAVA 1207

Qy         61 CDVGERREVVRLLE----GVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAA 116
               |  :   :  | :     :|   || |:  ||      | :: :| : : |:|  |:||
Db       1208 ADATDPAAMQTLFDRFGTELP---PLEGIYLAAFAGRPALLSEMTDDDVTTMFRPKLDAL 1264

Qy        117 LLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWG 176
               ||   :    :  |||||| :|:|||    :| | :| ||: |  ||  |||   : ||
Db       1265 ALLHRRSLKSPVRHFVLFSSVSGLLGSRWLAHYTATSAFLDSFAGARRTMGLPATVVDWG 1324

Qy        177 LW 178
               ||
Db       1325 LW 1326


RESULT 11
FAS_CHICK
ID   FAS_CHICK      STANDARD;      PRT;  2511 AA.
AC   P12276;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=White leghorn; TISSUE=Liver;
RX   MEDLINE=95031085; PubMed=7944406;
RA   Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Amino-terminal blocking group and sequence of the animal fatty acid
RT   synthase.";
RL   Arch. Biochem. Biophys. 314:45-49(1994).
RN   [2]
RP   SEQUENCE OF 75-1775 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89282777; PubMed=2734291;
RA   Holzer K.P., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of chicken liver fatty acid
RT   synthase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN   [3]
RP   SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89139426; PubMed=2917973;
RA   Chirala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,
RA   Wakil S.J.;
RT   "A novel cDNA extension procedure. Isolation of chicken fatty acid
RT   synthase cDNA clones.";
RL   J. Biol. Chem. 264:3750-3757(1989).
RN   [4]
RP   SEQUENCE OF 1752-2512 FROM N.A.
RX   MEDLINE=88320436; PubMed=2842766;
RA   Yuan Z., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of DNA complementary to chicken
RT   liver fatty acid synthase mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN   [5]
RP   SEQUENCE OF 2202-2512 FROM N.A.
RX   MEDLINE=89088152; PubMed=3207710;
RA   Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT   "Characterization of a genomic and cDNA clone coding for the
RT   thioesterase domain and 3' noncoding region of the chicken liver
RT   fatty acid synthase gene.";
RL   Biochemistry 27:7778-7785(1988).
RN   [6]
RP   SEQUENCE OF 2121-2209.
RX   MEDLINE=89192401; PubMed=2648999;
RA   Huang W.-Y., Stoops J.K., Wakil S.J.;
RT   "Complete amino acid sequence of chicken liver acyl carrier protein
RT   derived from the fatty acid synthase.";
RL   Arch. Biochem. Biophys. 270:92-98(1989).
RN   [7]
RP   SEQUENCE OF 2209-2508.
RC   STRAIN=White leghorn;
RX   MEDLINE=89088151; PubMed=3207709;
RA   Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Complete amino acid sequence of the thioesterase domain of chicken
RT   liver fatty acid synthase.";
RL   Biochemistry 27:7773-7777(1988).
RN   [8]
RP   SEQUENCE OF 667-674 AND 1698-1709.
RX   MEDLINE=89323081; PubMed=2751995;
RA   Chang S.I., Hammes G.G.;
RT   "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
RT   fluorescence resonance energy transfer in chicken liver fatty acid
RT   synthase.";
RL   Biochemistry 28:3781-3788(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
```

```
CC       POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04485; AAB46389.1; -.
DR   EMBL; J03860; AAA48767.1; -.
DR   EMBL; J02839; AAA82106.1; ALT_SEQ.
DR   PIR; A33918; XYCHFA.
DR   PIR; A32015; A32015.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;
KW   Alternative splicing; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   DOMAIN        1   ?411       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      427    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1638   1866       ENOYL REDUCTASE.
FT   DOMAIN     1867   2119       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2124   2180       ACYL CARRIER (ACP).
FT   DOMAIN     2209   2511       THIOESTERASE.
FT   MOD_RES       1      1       ACETYLATION.
FT   ACT_SITE    160    160       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    579    579       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   ACT_SITE    877    877       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   NP_BIND    1674   1691       NADP (ER).
FT   BINDING    1707   1707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1888   1903       NADP (KR).
FT   BINDING    2157   2157       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2308   2308       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2481   2481       THIOESTERASE (BY SIMILARITY).
FT   VARSPLIC   2348   2348       T -> TQCFSFSLF (IN ISOFORM 1).
FT   CONFLICT     77     78       QL -> PV (IN REF. 2).
FT   CONFLICT    116    116       L -> A (IN REF. 2).
FT   CONFLICT    675    675       R -> S (IN REF. 2).
FT   CONFLICT   1169   1169       K -> N (IN REF. 2).
FT   CONFLICT   1178   1178       A -> T (IN REF. 2).
FT   CONFLICT   1191   1191       R -> H (IN REF. 2).
FT   CONFLICT   1198   1198       P -> L (IN REF. 2).
FT   CONFLICT   1286   1287       DN -> ND (IN REF. 2).
FT   CONFLICT   1372   1372       K -> E (IN REF. 2).
FT   CONFLICT   1533   1533       C -> Y (IN REF. 2).
FT   CONFLICT   1577   1577       W -> R (IN REF. 2).
FT   CONFLICT   1685   1696       QAAIAIALSMGC -> ASSHCHRLEHGLA
FT                                (IN REF. 2; AAA48767).
FT   CONFLICT   1732   1732       Q -> E (IN REF. 2).
FT   CONFLICT   1745   1745       S -> N (IN REF. 2).
SQ   SEQUENCE   2511 AA;  274648 MW;  622039DAC8315D3F CRC64;

  Query Match           25.6%;  Score 225.5;  DB 1;  Length 2511;
  Best Local Similarity  36.2%;  Pred. No. 1.6e-09;
  Matches   64;  Conservative   26;  Mismatches   84;  Indels    3;  Gaps    2;

Qy          2 SVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVAC 61
              | ::||| |  |  :|: |    | : |:| || |           |  ||| :| :
Db       1888 SYIITGGLGGFGLELAQWLI-ERGAQKLVLTSRSGIRTGYQAKCVREWKALGIQVLVSTS 1946

Qy         62 DVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDE 121
              |||       |:|      |: |: : | || || | : |||    |   |    | ||
Db       1947 DVGTLEGTQLLIEEALKLGPVGGIFNLAVVLKDAMIENQTPELFWEVNKPKYSGTLHLDW 2006

Qy        122 LTRGM--ELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWG 176
               :||    :|  ||:||| :   |:||| ||  ||:|:: :  :|   ||||::: ||
Db       2007 VTRKKCPDLDYFVVFSSVSCGRGNAGQSNYGFANSAMERICEQRHHDGLPGLAVQWG 2063


RESULT 12
FAS_HUMAN
ID   FAS_HUMAN      STANDARD;      PRT;  2504 AA.
AC   P49327;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96004605; PubMed=7567999;
RA   Jayakumar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M.,
RA   Abu-Elheiga L., Chirala S.S., Wakil S.J.;
RT   "Human fatty acid synthase: properties and molecular cloning.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699(1995).
RN   [2]
RP   SEQUENCE OF 753-758 AND 1285-1297.
RX   MEDLINE=94294385; PubMed=8022791;
RA   Kuhajda F.P., Jenner K., Wood F.D., Hennigar R.A., Jacobs L.B.,
RA   Dick J.D., Pasternack G.R.;
RT   "Fatty acid synthesis: a potential selective target for
RT   antineoplastic therapy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383(1994).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND
CC       LIVER.
CC   -!- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY
CC       MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTETHEINE CONTENT OF
CC       THE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U26644; AAC50259.1; -.
DR   Genew; HGNC:3594; FASN.
DR   MIM; 600212; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      428    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1630   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2504       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    580    580       MALONYLTRANSFERASE (BY SIMILARITY).
FT   ACT_SITE    876    876       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   NP_BIND    1666   1683       NADP (ER).
FT   BINDING    1699   1699       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1880   1895       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2504 AA;  273100 MW;  8AAF9809B2338DFA CRC64;

  Query Match             24.3%;  Score 214.5;  DB 1;  Length 2504;
  Best Local Similarity   31.4%;  Pred. No. 1.1e-08;
  Matches   64;  Conservative   22;  Mismatches   61;  Indels   57;  Gaps    4;

Qy          2 SVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRG------------------------ 36
              | :: || |  |  :|: |    ||: |:| || |
Db       1880 SYIIAGGLGGFGLELAQWLI-QRGVQKLVLTSRSGIRTGYQAKQVRRWRRQGLQVQVSTS 1938

Qy         37 --PDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDD 94
                || || || | ||                          |: || : | || |
Db       1939 NISSLEGARGLIAEAAQLG----------------------PVGGVFNLAVVLRD 1971

Qy         95 ATIASLTPERLGTVFAAKVDAALLLDELTRGM--ELSAFVLFSSAAGILGSAGQGNYAAA 152
                : : |||    |   |    | || :||    ||  ||:||| :   |:||| ||  |
Db       1972 GLLENQTPEFFQDVCKPKYSGTLNLDRVTREACPELDYFVVFSSVSCGRGNAGQSNYGFA 2031

Qy        153 NAALDALAYRRRAAGLPGVSLAWG 176
              |:|:: :  :||  ||||::: ||
Db       2032 NSAMERICEKRRHEGLPGLAVQWG 2055


RESULT 13
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match           23.5%;  Score 207.5;  DB 1;  Length 2505;
  Best Local Similarity   34.8%;  Pred. No. 3.7e-08;
  Matches   63;  Conservative   25;  Mismatches   82;  Indels   11;  Gaps    4;

Qy          2 SVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRA----ELAALGAEVR 57
              | ::||| |  |  :|| |  : | : |:| || |       | :|    |    |  |
Db       1880 SYIITGGLGGFGLELARWLV-LRGAQRLVLTSRSGI----RTGYQAKHVREWRRQGIHVL 1934

Qy         58 IVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAAL 117
              :  :|       |:       |: || : | || || : : |||    |   | :  |
Db       1935 VSTSNVSSLEGARALIAEATKLGPVGGVFNLAMVLRDAMLENQTPELFQDVNKPKYNGTL 1994

Qy        118 LLDELTRGM--ELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAW 175
               ||  ||    ||  || ||| :  |:||| ||  ||: :: :  :||  ||||::: |
Db       1995 NLDRATREACPELDYFVAFSSVSCGRGNAGQSNYGFANSTMERICEQRRHDGLPGLAVQW 2054

Qy        176 G 176
               |
Db       2055 G 2055


RESULT 14
FAS_MOUSE
ID   FAS_MOUSE      STANDARD;      PRT;   838 AA.
AC   P19096;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]
DE   (Fragment).
GN   FASN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Liver;
RX   MEDLINE=89149781; PubMed=2920037;
RA   Paulauskis J.D., Sul H.S.;
RT   "Structure of mouse fatty acid synthase mRNA. Identification of the
RT   two NADPH binding sites.";
RL   Biochem. Biophys. Res. Commun. 158:690-695(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X13135; CAA31525.1; -.
DR   PIR; A32262; A32262.
DR   MGD; MGI:95485; Fasn.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; PARTIAL.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP.
FT   NON_TER       1      1
```

```
FT   DOMAIN       451    507       ACYL CARRIER (ACP).
FT   DOMAIN       535    838       THIOESTERASE.
FT   BINDING      484    484       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE     635    635       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE     808    808       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   838 AA;  91212 MW;  32C40EE5CA5AB841 CRC64;

  Query Match             21.1%;  Score 186.5;  DB 1;  Length 838;
  Best Local Similarity   33.1%;  Pred. No. 5e-07;
  Matches   60;  Conservative   26;  Mismatches   84;  Indels   11;  Gaps    4;

Qy          2 SVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRA----ELAALGAEVR 57
              | ::||| |   |  :|| |  : | : |:| || |       | :|    |    | :|
Db        213 SYIITGGLGGFGLELARWLV-LRGAQRLVLTSRSGI----RTGYQAKHIREWRRQGIQVL 267

Qy         58 IVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAAL 117
              :  :|       |:       |: || :    | || : : |||    |   | : |
Db        268 VSTSNVSSLEGARALIAEATKLGPVGGVFNLGHGLRDAMLENQTPELFQDVNKPKYNGNL 327

Qy        118 LLDELTRGM--ELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAW 175
               ||  ||    ||  || ||| :   |:||| ||  ||: :: :  :||   |||::: |
Db        328 NLDRATREACPELDYFVAFSSVSCGRGNAGQTNYCFANSTMERICEQRRHDVLPGLAVQW 387

Qy        176 G 176
              |
Db        388 G 388


RESULT 15
PKSM_BACSU
ID   PKSM_BACSU     STANDARD;      PRT;  4273 AA.
AC   P40872; O31781;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksM.
GN   PKSM OR PKSY.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE OF 1-1763 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Tognoni A., Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99113; CAB13603.1; -.
DR   EMBL; Z35133; CAA84505.1; -.
DR   SubtiList; BG10931; pksM.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001601; Methyltransf.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN       295    364       ACYL CARRIER (ACP) 1.
FT   DOMAIN       396    834       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      2190   2258       ACYL CARRIER (ACP) 2.
FT   DOMAIN      2322   2737       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN      3532   3947       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN      3410   3483       ACYL CARRIER (ACP) 3.
FT   DOMAIN      4140   4209       ACYL CARRIER (ACP) 4.
FT   BINDING      327    327       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING     2222   2222       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    2476   2476       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING     3446   3446       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    3690   3690       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING     4172   4172       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT     103    103       E -> V (IN REF. 2).
FT   CONFLICT     276    276       Q -> E (IN REF. 2).
FT   CONFLICT     289    289       T -> S (IN REF. 2).
SQ   SEQUENCE   4273 AA;  477459 MW;  3BBFCF1A250AEB5A CRC64;

  Query Match             19.2%;  Score 169;  DB 1;  Length 4273;
  Best Local Similarity   29.0%;  Pred. No. 5e-05;
  Matches   53;  Conservative   27;  Mismatches   93;  Indels   10;  Gaps    4;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |: |:||| | ||   |:|||         |::: | |      :    ||  || |
Db       3145 GTYLITGGVGGLGYLFAKHLAKNYAAN--LILTGRSPFNDEKQKQIKELKDLGGERMYAE 3202

Qy         61 CDVGERREVVRLLEGVPAG----CPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAA 116
               ||    : : : : |  |      : ||:||||:  |: |    |    :   |::
Db       3203 ADVS---DPIAMGDCVKRGKDRFGAINGVIHAAGIESDSAIFDKKIESFQRIIEPKINGT 3259

Qy        117 LLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALA-YRRRAAGLPGVSLAW 175
              : |||  :  :|     |||:: :||  |  :||  |    | | ||          : |
Db       3260 IALDEWLKNEDLDFMCYFSSSSAVLGDFGCCDYAIGNRFQMAYAQYRNELHNGKTFVINW 3319
```

Qy 176 GLW 178
:|
Db 3320 PVW 3322

Search completed: June 17, 2003, 13:02:40
Job time : 2.94115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 3.97473 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_3537_3714
Perfect score: 882
Sequence: 1 GSVLVTGGTGVLGAAVARHL.....LAYRRRAAGLPGVSLAWGLW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	561	63.6	4151	2	G70944	probable polyketid
2	547	62.0	3739	2	T17410	polyketide synthas
3	546	61.9	4613	2	T17409	polyketide synthas
4	520	59.0	4735	2	T17463	rifamycin polyketi
5	503.5	57.1	3413	2	T17467	rifamycin polyketi
6	495.5	56.2	1728	2	T17466	rifamycin polyketi
7	492	55.8	5069	2	T17464	rifamycin polyketi
8	488.5	55.4	10223	2	T30225	polyketide synthas
9	482.5	54.7	1616	2	G70668	polyketide synthas
10	481.5	54.6	1937	2	T03224	probable polyketid
11	481	54.5	6260	2	T30228	polyketide synthas
12	478	54.2	2126	2	H70621	probable polyketid
13	477	54.1	8563	2	T30226	polyketide synthas
14	476.5	54.0	2723	2	T03221	probable polyketid
15	475	53.9	2124	2	T28658	polyketide synthas
16	468.5	53.1	502	2	A70985	probable polyketid
17	467.5	53.0	2103	2	G86925	probable polyketid
18	456.5	51.8	1762	2	T03222	probable polyketid
19	456.5	51.8	2100	2	T03223	probable polyketid
20	450.5	51.1	3573	2	S23070	erythronolide synt
21	433.5	49.1	1763	2	T17465	rifamycin polyketi
22	427	48.4	3172	2	S22012	erythronolide synt
23	426.5	48.4	3491	2	T43231	probable 6-deoxyer
24	425.5	48.2	1562	2	T17411	polyketide synthas
25	403.5	45.7	6420	2	T30283	polyketide synthas
26	402	45.6	3178	2	S13595	6-deoxyerythronoli
27	392	44.4	3519	2	S43048	polyketide synthas
28	374.5	42.5	7576	2	T17428	FK506 polyketide s
29	352.5	40.0	2518	2	AI2140	polyketide synthas
30	347	39.3	1198	2	T28678	polyketide synthas
31	338	38.3	1570	2	AC2012	hypothetical prote
32	330.5	37.5	2126	2	E70522	probable polyketid
33	328.5	37.2	1582	2	E70876	probable polyketid
34	324.5	36.8	2478	2	AH2140	polyketide synthas
35	302.5	34.3	2111	2	A70668	mycocerosate synth
36	299.5	34.0	2108	2	H70819	probable polyketid
37	295.5	33.5	2110	2	B44110	mycocerosate synth
38	295.5	33.5	2116	2	C86926	probable mycoceros
39	287.5	32.6	2188	2	A70984	probable polyketid
40	285.5	32.4	2118	2	S72705	mycocerosate synth
41	271.5	30.8	2201	2	S73014	polyketide synthas
42	267	30.3	1827	2	B70984	probable polyketid
43	266	30.2	1774	2	T17421	polyketide synthas
44	265	30.0	958	2	S73012	polyketide synthas
45	265	30.0	1871	2	A87204	polyketide synthas

ALIGNMENTS

RESULT 1
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

Query Match 63.6%; Score 561; DB 2; Length 4151;
Best Local Similarity 60.7%; Pred. No. 9.7e-36;
Matches 108; Conservative 31; Mismatches 39; Indels 0; Gaps 0;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||: |:|||||:    |||:|:||||||||||||  | ||||| ||:|::|| .
Db       1680 GTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVAELAAAGAQVQVVA 1739

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
                ||  :|  : :::  :|   ||:||:| || |||| : ||||:|:  |  :|||||  |
Db       1740 CDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRVDVVLRSKVDAAWHLH 1799

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| :::|||||||| ||::||:|| ||||||: |||||  ||| ||| :|| ||||
Db       1800 ELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALAAHRRAHGLPAISLGWGLW 1857
```

RESULT 2
T17410

polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

Query Match 62.0%; Score 547; DB 2; Length 3739;
Best Local Similarity 62.8%; Pred. No. 1.1e-34;
Matches 113; Conservative 19; Mismatches 46; Indels 2; Gaps 1;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy 121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456

RESULT 3
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

Query Match 61.9%; Score 546; DB 2; Length 4613;
Best Local Similarity 62.8%; Pred. No. 1.6e-34;
Matches 113; Conservative 20; Mismatches 45; Indels 2; Gaps 1;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy 121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279

RESULT 4
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

Query Match 59.0%; Score 520; DB 2; Length 4735;
Best Local Similarity 61.2%; Pred. No. 1.8e-32;
Matches 109; Conservative 20; Mismatches 45; Indels 4; Gaps 2;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 1829 GTVLITGGTGTLGALTARHLVTAHGVRHLVLVSRRG-EAP---ELQEELTALGASVAIAA 1884

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 1885 CDVADRAQLEAVLRAIPAEHPLTAVIHTAGVLDDGVVTELTPDRLATVRRPKVDAARLLD 1944

Qy 121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 1945 ELTREADLAAFVLFSSAAGVLGNPGQAGYAAANAELDALARQRNSLDLPAVSIAWGYW 2002

RESULT 5
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

Query Match 57.1%; Score 503.5; DB 2; Length 3413;
Best Local Similarity 63.1%; Pred. No. 2.6e-31;
Matches 113; Conservative 14; Mismatches 45; Indels 7; Gaps 3;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 1338 GTVLIS-GAGSLGALTARHLVARHGVRRLVLVSRRGPDADGMAELTAELIAQGAEVAVVA 1396

Qy 61 CDVGERREV-VRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119

```
Db       1397 CDLADRDQVRVLLAEHRP-----NAVVHTAGVLDDGVFESLTRERLAKVFAPKVTAANHL 1451

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178

Db       1452 DELTRELDLRAFVVFSSASGVFGSAGQGNYAAANAYLDAVVANRRAAGLPGTSLAWGLW 1510


RESULT 6
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17466
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17466
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1728 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1573-1644/Domain: acyl carrier protein homology <ACP>

  Query Match             56.2%;  Score 495.5;  DB 2;  Length 1728;
  Best Local Similarity   60.3%;  Pred. No. 5.5e-31;
  Matches  108;  Conservative   19;  Mismatches   45;  Indels    7;  Gaps    3;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60

Db       1300 GTVLVS-GAGVLGEIVARHLVTRHGVRKLVLASRRGLDADGAKDLVTDLTGEGADVSVVA 1358

Qy         61 CDVGERREVVRLL-EGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119

Db       1359 CDLADRNQVAALLADHRPA-----SVIHTAGVLDDGVIGTLTPERLAKVFAPKVDAVRHL 1413

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178

Db       1414 DELTRDLDLDAFVVFSSGSGVFGSPGQGNYAAANAFLDAAMASRRAAGLPGLSLAWGLW 1472


RESULT 7
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             55.8%;  Score 492;  DB 2;  Length 5069;
  Best Local Similarity   59.0%;  Pred. No. 3.1e-30;
  Matches  105;  Conservative   22;  Mismatches   47;  Indels    4;  Gaps    2;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60

Db       4666 GTVLVS-GAGSLGTLVARHLVTRHGVRRVVLASRQGRDAEGAQDLITELTGEGADVSFVA 4724

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120

Db       4725 CDVSDRDQVAALLAGLP---DLTGVVHTAGVFEDGVIEALTPDQLANVYAAKVTAAMHLD 4781

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178

Db       4782 ELTRDRDLGAFVVFSSVAGVMGGGQGPYAAANAFLDAAMASRQAAGLPGLSLAWGLW 4839


RESULT 8
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staur
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match             55.4%;  Score 488.5;  DB 2;  Length 10223;
  Best Local Similarity   58.7%;  Pred. No. 1.2e-29;
  Matches  105;  Conservative   20;  Mismatches   49;  Indels    5;  Gaps    2;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60

Db       9800 GTVLITGGSGVLAGIAARHLVAERGVRHLLLLSRSAPD----EALISELAELGAAVVDTA 9855

Qy         61 -CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119

Db       9856 VCDVSDRAGLARVLAGVSPDHPLTAVIHTAGVLDDGVVESLTARRLDTVLRPKADGAWNL 9915

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178

Db       9916 HELTRDIDLAAFVMYSSAAGVLGSAGQGNYAVANAFVDALAEQRRAEGLPALALAWGLW 9974


RESULT 9
G70668
polyketide synthase pks1 - Mycobacterium tuberculosis (strain H37RV)
N;Alternate names: polyketide synthase pks002c
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70668; S73075
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
```

A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70668
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1616 <COL>
A;Cross-references: GB:Z83858; GB:AL123456; NID:g3261675; PIDN:CAB06103.1; PID:g1781167
A;Experimental source: strain H37Rv
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A;Description: Mycobacterium tuberculosis cosmid tbc2.
A;Reference number: S73053
A;Accession: S73075
A;Molecule type: DNA
A;Residues: 'MWGARWPVARCLSTEL',48-917,'RSEARSRRGCGMSGRAAPPAPPIGPPIGIP','PMPPIIPVEPAGPVFN
A;Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50928.1; PID:g560507
C;Genetics:
A;Gene: pks1
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase pks1; acyl carrie
ology
C;Keywords: carrier protein
F;84-362/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1228-1410/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1514-1585/Domain: acyl carrier protein homology <ACP1>

Query Match 54.7%; Score 482.5; DB 2; Length 1616;
Best Local Similarity 56.4%; Pred. No. 5.3e-30;
Matches 101; Conservative 23; Mismatches 54; Indels 1; Gaps 1;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 1228 GTVVITGATGAVGGVLARHLVGAYGVRHLVLASRRGDRAEGAAELAADLTEAGAKVQVVA 1287

Qy 61 CDVGERREVVRLLEGVPAGC-PLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
Db 1288 CDVADRAAVAGLFAQLSREYPPVRGVIHAAGVLDDAVITSLTPDRIDTVLRAKVDAAWNL 1347

Qy 120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 1348 HQATSDLDLSMFALCSSIAATVGSPGQGNYSAANAFLDGLAAHRQAAGLAGISLAWGLW 1406

RESULT 10
T03224
probable polyketide synthase module 4 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03224
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03224
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1937 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38064.1; PID:g2624951
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;549-827/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

Query Match 54.6%; Score 481.5; DB 2; Length 1937;
Best Local Similarity 57.3%; Pred. No. 7.7e-30;
Matches 102; Conservative 19; Mismatches 54; Indels 3; Gaps 1;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 1319 GTILITGGTGSLGSLVARHLVTEHGVRDLVLVSRQGPDADSATELAAELQQHGARVRTLS 1378

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 1379 CDLTDRTALATLIDEIG---PLTGVVHTAGALADATVEHLDAEALATTFAPKANTAWWLH 1435

Qy 121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 1436 ELTQDHDLTLFALFSSVAGILGSPGQANYAAANTFLDALATHRRGLGLPGTSLAWGSW 1493

RESULT 11
T30228
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30228
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30228
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6260 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987102; PIDN:CAA60462.1
C;Genetics:
A;Gene: rapC
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1536-1607/Domain: acyl carrier protein homology <ACP1>
F;1651-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3123-3194/Domain: acyl carrier protein homology <ACP2>
F;3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3725-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5114-5185/Domain: acyl carrier protein homology <ACP3>
F;5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6135-6206/Domain: acyl carrier protein homology <ACP4>

Query Match 54.5%; Score 481; DB 2; Length 6260;
Best Local Similarity 56.7%; Pred. No. 2.8e-29;
Matches 101; Conservative 21; Mismatches 52; Indels 4; Gaps 1;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 1270 GTVLITGGSGVLAGIAARHLVTERGVRHLLLLSRSAPD----EALIGELGELGARVETAA 1325

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 1326 CDVSDPAALTQVLAGVSPEHPLTAVIHTAGVVDDGVVESLTVQRLETVLRPKADGAWNLH 1385

Qy 121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 1386 ELTRDADLAAFVMYSSAAGVLGSAGQGNYAAANAFLDALAEQRHAEGLPALAVAWGLW 1443

RESULT 12
H70621
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70621
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2126 <COL>

```
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06632.1; PID:g1839013
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks7
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;563-845/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1694-1875/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1976-2051/Domain: acyl carrier protein homology <ACP2>

  Query Match             54.2%;  Score 478;  DB 2;  Length 2126;
  Best Local Similarity   57.7%;  Pred. No. 1.6e-29;
  Matches  105;  Conservative   22;  Mismatches   47;  Indels    8;  Gaps    2;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAAL----GAEV 56
              |:|:||||||: |:||| ||    || :|:|||| |  |  |    ||:|||    ||:|
Db       1694 GTVVVTGGTGMAGSAVATHLVRRHGVANLVLVSRSGEQADRA----AEVAALLREGGAQV 1749

Qy         57 RIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAA 116
               :|:||| :|  :  || |:    || || ||||||||| |  |||:|: ||  |||| |
Db       1750 AVVSCDVADRDALAALLAGLDPRYPLKGVFHAAGVLDDAVITGLTPDRVDTVLRAKVDGA 1809

Qy        117 LLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWG 176
              | |||  |:|||||:||| |||:|:  |||||||| || |   ||: || |:|:|||
Db       1810 WNLHELTEDMDLSAFVVFSSMAGIVGTPAQGNYAAANAFLDGLVAYRRSRGLAGLSVAWG 1869

Qy        177 LW 178
              ||
Db       1870 LW 1871


RESULT 13
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             54.1%;  Score 477;  DB 2;  Length 8563;
  Best Local Similarity   55.1%;  Pred. No. 7.9e-29;
  Matches   98;  Conservative   25;  Mismatches   51;  Indels    4;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:|:|:||:|||   :|||||   | | |||:||  ||    : |  ||| ||| |
Db       2931 GTILITGGSGVLAGILARHLAAEHGARHLLLLSRTTPD----QALIKELAELGAHVDTAT 2986

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  : |:|  ||     ||| |:| || |||  : ||| ::| ||   | | |  |
Db       2987 CDVSDRAGLARVLAGVSPEHPLTAVIHTAGALDDGVVESLTTQQLDTVLRPKADGAWHLH 3046

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||:  :|:|||::|||||:|||||||||||||| :|||| :||  ||| :::|||||
Db       3047 ELTQNTDLAAFVMYSSAAGVLGSAGQGNYAAANAFVDALAEQRRGEGLPALAVAWGLW 3104


RESULT 14
T03221
probable polyketide synthase module 1 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 17-Nov-2000
C;Accession: T03221
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03221
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2723 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38061.1; PID:g2624948
A;Experimental source: ATCC 29253
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-504/Domain: acetate-CoA ligase homology <ACL>
F;881-949/Domain: acyl carrier protein homology <ACP1>
F;996-1397/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;1509-1794/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;2568-2639/Domain: acyl carrier protein homology <ACP2>

  Query Match             54.0%;  Score 476.5;  DB 2;  Length 2723;
  Best Local Similarity   56.7%;  Pred. No. 2.7e-29;
  Matches  101;  Conservative   22;  Mismatches   52;  Indels    3;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||: :||||    |||||:||||:|||||||  | |||  |||:||: :
Db       2292 GTVLITGGTGSLGSLIARHLVTEHGVRDLVLVSRQGPDAPGATELIAELQQLGAQVRVRS 2351

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|  :  |:: |    |  |||| || | ||| :|  : | |  |||  ||  |
Db       2352 CDTTDRAALAALVDEVG---PFAGVVHTAGALADATVDNLDADALATTIAAKAHAAWWLH 2408

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              | |:   |: |::||| | :|||||| ||||||: |||||  ||  |||| ||||| |
Db       2409 EHTQDHGLTLFMVFSSLASVLGSAGQANYAAANSFLDALATHRRTLGLPGTSLAWGSW 2466


RESULT 15
T28658
polyketide synthase - Sorangium cellulosum (fragment)
C;Species: Sorangium cellulosum
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C;Accession: T28658
R;Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J.J.; Ligon, J.M.
J. Bacteriol. 177, 3673-3679, 1995
A;Title: A Sorangium cellulosum (myxobacterium) gene cluster for the biosynthesis of the
inomycetes.
A;Reference number: Z20499; MUID:95325306; PMID:7601830
A;Accession: T28658
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2124 <SCH>
A;Cross-references: EMBL:U24241; NID:g868150; PID:g868151; PIDN:AAA79984.1
A;Experimental source: strain So ce26
C;Function:
A;Description: involved in soraphen A biosynthesis
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;335-406/Domain: acyl carrier protein homology <ACP>
```

F;370/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 53.9%; Score 475; DB 2; Length 2124;
Best Local Similarity 55.1%; Pred. No. 2.7e-29;
Matches 98; Conservative 27; Mismatches 53; Indels 0; Gaps 0;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||| ||| |      : ||| ||:|  ||||: ||:|| |||| | : |
Db       1819 GTVLITGGTGTLGALVARRLVVNHDAKHLLLTSRQGASAPGADVLRSELEALGASVTLAA 1878

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| : | :  ||: :|:  |:  ||||| |||   : ::: ||:  ||| |:|||  |
Db       1879 CDVADPRALKDLLDNIPSAHPVAAVVHAASVLDGDLLGAMSLERIDRVFAPKIDAAWHLH 1938

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              :||:   |:||:|||| ||:|||:|  |||||:|.|||||: ||| |||  |||| |
Db       1939 QLTQDKPLAAFILFSSVAGVLGSSGHSNYAAASAFLDALAHHRRAQGLPASSLAWSHW 1996
```

Search completed: June 17, 2003, 13:12:25
Job time : 4.97473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 5.54613 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_3537_3714
Perfect score: 882
Sequence: 1 GSVLVTGGTGVLGAAVARHL.....LAYRRRAAGLPGVSLAWGLW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	547	62.0	3739	9	US-09-860-846-33	Sequence 33, Appl
2	547	62.0	3739	9	US-09-988-384B-33	Sequence 33, Appl
3	547	62.0	3739	9	US-09-836-821-33	Sequence 33, Appl
4	547	62.0	3739	9	US-09-793-708-2	Sequence 2, Appli
5	547	62.0	3739	10	US-09-861-289-33	Sequence 33, Appl
6	547	62.0	11877	9	US-09-860-846-6	Sequence 6, Appli
7	547	62.0	11877	9	US-09-836-821-6	Sequence 6, Appli
8	547	62.0	11877	10	US-09-861-289-6	Sequence 6, Appli
9	547	62.0	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	546	61.9	4551	9	US-09-793-708-1	Sequence 1, Appli
11	546	61.9	4613	9	US-09-860-846-31	Sequence 31, Appl
12	546	61.9	4613	9	US-09-988-384B-31	Sequence 31, Appl
13	546	61.9	4613	9	US-09-836-821-31	Sequence 31, Appl
14	546	61.9	4613	10	US-09-861-289-31	Sequence 31, Appl
15	530	60.1	5215	9	US-09-860-846-2	Sequence 2, Appli
16	530	60.1	5215	9	US-09-988-384B-2	Sequence 2, Appli
17	530	60.1	5215	9	US-09-836-821-2	Sequence 2, Appli
18	530	60.1	5215	10	US-09-861-289-2	Sequence 2, Appli
19	482.5	54.7	1616	9	US-09-712-363-262	Sequence 262, App
20	449.5	51.0	3816	9	US-09-808-880-3	Sequence 3, Appli
21	440.5	49.9	7257	9	US-10-014-717-5	Sequence 5, Appli
22	428	48.5	4150	9	US-09-808-880-2	Sequence 2, Appli
23	425.5	48.2	1562	9	US-09-860-846-35	Sequence 35, Appl
24	425.5	48.2	1562	9	US-09-988-384B-35	Sequence 35, Appl
25	425.5	48.2	1562	9	US-09-836-821-35	Sequence 35, Appl
26	425.5	48.2	1562	9	US-09-793-708-3	Sequence 3, Appli
27	425.5	48.2	1562	10	US-09-861-289-35	Sequence 35, Appl
28	395	44.8	184	9	US-09-957-483-33	Sequence 33, Appl
29	392	44.4	3519	9	US-09-808-880-4	Sequence 4, Appli
30	391	44.3	2439	9	US-10-014-717-7	Sequence 7, Appli
31	381	43.2	184	9	US-09-957-483-35	Sequence 35, Appl
32	380	43.1	1832	9	US-10-014-717-4	Sequence 4, Appli
33	336.5	38.2	3798	9	US-10-014-717-6	Sequence 6, Appli
34	267	30.3	1827	9	US-09-712-363-261	Sequence 261, App
35	241.5	27.4	2563	9	US-09-836-705-46	Sequence 46, Appl
36	214.5	24.3	2509	9	US-10-237-271-1	Sequence 1, Appli
37	196	22.2	3032	9	US-09-836-705-44	Sequence 44, Appl
38	177	20.1	711	9	US-10-124-800-18	Sequence 18, Appl
39	177	20.1	1481	9	US-10-331-061-70	Sequence 70, Appl
40	177	20.1	2910	9	US-10-124-800-2	Sequence 2, Appli
41	165.5	18.8	2756	9	US-10-331-061-7	Sequence 7, Appli
42	132	15.0	247	10	US-09-815-242-11893	Sequence 11893, A
43	122.5	13.9	243	10	US-09-815-242-13360	Sequence 13360, A
44	122.5	13.9	243	10	US-09-815-242-13581	Sequence 13581, A
45	116.5	13.2	246	9	US-09-738-626-5141	Sequence 5141, Ap

ALIGNMENTS

RESULT 1
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33

Query Match 62.0%; Score 547; DB 9; Length 3739;
Best Local Similarity 62.8%; Pred. No. 5.5e-40;
Matches 113; Conservative 19; Mismatches 46; Indels 2; Gaps 1;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| | |||  |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
                ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
                |||   | :|:|||:||||| : | |||| |||||| ||||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456
```

RESULT 2
US-09-988-384B-33

```
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             62.0%;  Score 547;  DB 9;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 5.5e-40;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| | |||  |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 3
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             62.0%;  Score 547;  DB 9;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 5.5e-40;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| | |||  |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 4
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-2

  Query Match             62.0%;  Score 547;  DB 9;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 5.5e-40;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| | |||  |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 5
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             62.0%;  Score 547;  DB 10;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 5.5e-40;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||:|  ||| ||||||||  |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||  : | |||| ||||||  |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match             62.0%;  Score 547;  DB 9;  Length 11877;
  Best Local Similarity   62.8%;  Pred. No. 2.2e-39;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||:|  ||| ||||||||  |||||  |  || ||||:| : |
Db       8226 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 8285

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       8286 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 8345

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||  : | |||| ||||||  |||||:||| |||| :|| ||||
Db       8346 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 8405


RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-6

  Query Match             62.0%;  Score 547;  DB 9;  Length 11877;
  Best Local Similarity   62.8%;  Pred. No. 2.2e-39;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||:|  ||| ||||||||  |||||  |  || ||||:| : |
Db       8226 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 8285

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       8286 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 8345

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||  : | |||| ||||||  |||||:||| |||| :|| ||||
Db       8346 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 8405


RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             62.0%;  Score 547;  DB 10;  Length 11877;
  Best Local Similarity   62.8%;  Pred. No. 2.2e-39;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||:|  ||| ||||||||  |||||  |  || ||||:| : |
Db       8226 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 8285

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       8286 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 8345

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||  : | |||| ||||||  |||||:||| |||| :|| ||||
Db       8346 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 8405


RESULT 9
US-09-988-384B-6
```

```
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             62.0%;  Score 547;  DB 9;  Length 12199;
  Best Local Similarity   62.8%;  Pred. No. 2.3e-39;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db       8548 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 8607

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :   |  ||||| |||
Db       8608 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 8667

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||    | :|:|||:|||||  : | |||| |||||| |||||:||| |||| :|| ||||
Db       8668 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 8727


RESULT 10
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-1

  Query Match             61.9%;  Score 546;  DB 9;  Length 4551;
  Best Local Similarity   62.8%;  Pred. No. 8.5e-40;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||: |  ||| |||||||| |||||: |  || ||||:| : |
Db       4038 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4097

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :   |  ||||| |||
Db       4098 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4157

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||     :|:|||:|||||  : | |||| |||||| |||||:|||||||| :|| ||||
Db       4158 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4217


RESULT 11
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-31

  Query Match             61.9%;  Score 546;  DB 9;  Length 4613;
  Best Local Similarity   62.8%;  Pred. No. 8.7e-40;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||||  ||  ||||: |  ||| |||||||| |||||: |  || ||||:| : |
Db       4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :   |  ||||| |||
Db       4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||     :|:|||:|||||  : | |||| |||||| |||||:|||||||| :|| ||||
Db       4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279


RESULT 12
US-09-988-384B-31
; Sequence 31, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
```

```
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-31

  Query Match             61.9%;  Score 546;  DB 9;  Length 4613;
  Best Local Similarity   62.8%;  Pred. No. 8.7e-40;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| |  ||| |||||||| |||||: |  || ||||:| : |
Db       4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||     :|:|||:||||| : | |||| |||||| |||||:|||||||| :|| ||||
Db       4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279


RESULT 13
US-09-836-821-31
; Sequence 31, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-31

  Query Match             61.9%;  Score 546;  DB 9;  Length 4613;
  Best Local Similarity   62.8%;  Pred. No. 8.7e-40;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| |  ||| |||||||| |||||: |  || ||||:| : |
Db       4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| |\| |||| | |: |:| | :  |   ||||| |||
Db       4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||     :|:|||:||||| : | |||| |||||| |||||:|||||||| :|| ||||
Db       4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279


RESULT 14
US-09-861-289-31
; Sequence 31, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-31

  Query Match             61.9%;  Score 546;  DB 10;  Length 4613;
  Best Local Similarity   62.8%;  Pred. No. 8.7e-40;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||| |  ||| |||||||| |||||: |  || ||||:| : |
Db       4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||     :|:|||:||||| : | |||| |||||| |||||:|||||||| :|| ||||
Db       4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279


RESULT 15
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-2

  Query Match             60.1%;  Score 530;  DB 9;  Length 5215;
  Best Local Similarity   62.9%;  Pred. No. 2.7e-38;
  Matches  112;  Conservative   14;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||| :|||    ||| |||||||| |||||  |  :|| |||||   |
Db       2440 GTVLVTGGTGGLGALLARHLVERHGVRRLLLVSRRGADAPGAADLGEDLAGLGAEVAFAA 2499

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              |  :|  : | :  |||  ||| ||| |||:|||| :||||||  |   |||||  |
Db       2500 ADAADRESLARAIATVPAEHPLTAVVHTAGVVDDATVEALTPERLDAVLRPKVDAAWNLH 2559

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||: : | ||||||| :||:|:||| ||||||  |||||  | | ||   |||||||
Db       2560 ELTKDLRLDAFVLFSSVSGIVGTAGQANYAAANTGLDALAAHRAATGLAATSLAWGLW 2617


Search completed: June 17, 2003, 13:46:47
```

Job time : 6.54613 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 3.23524 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_3537_3714
Perfect score: 882
Sequence: 1 GSVLVTGGTGVLGAAVARHL.....LAYRRRAAGLPGVSLAWGLW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	568	64.4	3724	2	US-08-804-227C-10	Sequence 10, Appl
2	568	64.4	3724	2	US-08-804-198-4	Sequence 4, Appli
3	547	62.0	3739	3	US-09-320-878-2	Sequence 2, Appli
4	547	62.0	3739	4	US-09-105-537-33	Sequence 33, Appl
5	547	62.0	11877	4	US-09-105-537-6	Sequence 6, Appli
6	546	61.9	4551	3	US-09-320-878-1	Sequence 1, Appli
7	546	61.9	4613	4	US-09-105-537-31	Sequence 31, Appl
8	534	60.5	4472	2	US-08-804-227C-2	Sequence 2, Appli
9	530	60.1	5215	4	US-09-105-537-2	Sequence 2, Appli
10	520	59.0	5588	4	US-09-036-987A-6	Sequence 6, Appli
11	520	59.0	5588	4	US-09-370-700-6	Sequence 6, Appli
12	518	58.7	1864	2	US-08-804-227C-3	Sequence 3, Appli
13	512.5	58.1	6095	4	US-09-144-085-2	Sequence 2, Appli
14	509.5	57.8	5087	4	US-09-144-085-1	Sequence 1, Appli
15	504	57.1	4928	4	US-09-036-987A-5	Sequence 5, Appli
16	504	57.1	4928	4	US-09-370-700-5	Sequence 5, Appli
17	500	56.7	4545	2	US-08-804-227C-14	Sequence 14, Appl
18	500	56.7	4550	2	US-08-804-227C-8	Sequence 8, Appli
19	500	56.7	4550	2	US-08-804-198-2	Sequence 2, Appli
20	483.5	54.8	1841	2	US-08-804-227C-6	Sequence 6, Appli
21	482	54.6	3729	2	US-08-804-227C-4	Sequence 4, Appli
22	477	54.1	1996	2	US-08-804-227C-9	Sequence 9, Appli
23	477	54.1	1996	2	US-08-804-198-3	Sequence 3, Appli
24	475	53.9	2152	4	US-09-036-987A-3	Sequence 3, Appli
25	475	53.9	2152	4	US-09-370-700-3	Sequence 3, Appli
26	475	53.9	3567	2	US-07-642-734C-4	Sequence 4, Appli
27	475	53.9	3567	3	US-08-439-009A-4	Sequence 4, Appli
28	463.5	52.6	3170	4	US-09-036-987A-4	Sequence 4, Appli
29	463.5	52.6	3170	4	US-09-370-700-4	Sequence 4, Appli
30	449.5	51.0	3816	4	US-09-428-517-3	Sequence 3, Appli
31	440.5	49.9	7257	3	US-09-335-409-5	Sequence 5, Appli
32	440.5	49.9	7257	4	US-09-568-102-5	Sequence 5, Appli
33	440.5	49.9	7257	4	US-09-567-969-5	Sequence 5, Appli
34	440.5	49.9	7257	4	US-09-568-480-5	Sequence 5, Appli
35	440.5	49.9	7257	4	US-09-568-486-5	Sequence 5, Appli
36	440.5	49.9	7257	4	US-09-568-472-5	Sequence 5, Appli
37	440.5	49.9	7257	4	US-09-567-899-5	Sequence 5, Appli
38	438	49.7	3170	2	US-07-642-734C-5	Sequence 5, Appli
39	438	49.7	3170	3	US-08-439-009A-5	Sequence 5, Appli
40	431	48.9	159	3	US-09-010-809-1	Sequence 1, Appli
41	428.5	48.6	2595	4	US-09-036-987A-2	Sequence 2, Appli
42	428.5	48.6	2595	4	US-09-370-700-2	Sequence 2, Appli
43	428	48.5	4150	4	US-09-428-517-2	Sequence 2, Appli
44	426.5	48.4	3491	2	US-07-642-734C-2	Sequence 2, Appli
45	426.5	48.4	3491	3	US-08-439-009A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3724 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-10

Query Match 64.4%; Score 568; DB 2; Length 3724;
Best Local Similarity 65.2%; Pred. No. 6e-54;
Matches 116; Conservative 19; Mismatches 43; Indels 0; Gaps 0;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:|||||||| ||: :|||||   ||| ||: :|||||  ||  | |:||||||   :||
Db       3268 GTVLVTGGTGALGSRIARHLASRHGVRRLLIAARRGPDGEGAAELVADLAALGASATVVA 3327

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
```

```
                ||| :   |  || |:||  ||| |||: |||||  :  |||||:  |   ||:||: ||
Db       3328 CDVSDADAVRGLLAGIPADHPLTAVVHSTGVLDDGVLPGLTPERMRRVLRPKVEAAVHLD 3387

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::|||||||||:||:|||  |||||||| |||| ||| ||| ||||||||
Db       3388 ELTRDLDLSAFVLFSSSAGLLGSPAQGNYAAANATLDALAARRRSLGLPSVSLAWGLW 3445


RESULT 2
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-4

  Query Match             64.4%;  Score 568;  DB 2;  Length 3724;
  Best Local Similarity   65.2%;  Pred. No. 6e-54;
  Matches  116;  Conservative   19;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:|||||||| ||: :|||||   ||| ||: :|||||  ||  | |:||||||   :||
Db       3268 GTVLVTGGTGALGSRIARHLASRHGVRRLLIAARRGPDGEGAAELVADLAALGASATVVA 3327

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :   |  || |:||  ||| |||: |||||  :  |||||:  |   ||:||: ||
Db       3328 CDVSDADAVRGLLAGIPADHPLTAVVHSTGVLDDGVLPGLTPERMRRVLRPKVEAAVHLD 3387

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::|||||||||:||:|||  |||||||| |||| ||| ||| ||||||||
Db       3388 ELTRDLDLSAFVLFSSSAGLLGSPAQGNYAAANATLDALAARRRSLGLPSVSLAWGLW 3445


RESULT 3
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-2

  Query Match             62.0%;  Score 547;  DB 3;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 1.4e-51;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   |||||  |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:|||||  :  | |||| ||||||  |||||:||| ||||  :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 4
US-09-105-537-33
; Sequence 33, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-33

  Query Match             62.0%;  Score 547;  DB 4;  Length 3739;
  Best Local Similarity   62.8%;  Pred. No. 1.4e-51;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336
```

```
Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db        3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy         121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               |||   | :|:|||:||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db        3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456


RESULT 5
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             62.0%;  Score 547;  DB 4;  Length 11877;
  Best Local Similarity   62.8%;  Pred. No. 7.1e-51;
  Matches  113;  Conservative   19;  Mismatches   46;  Indels    2;  Gaps    1;

Qy           1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
               |:||:||||| ||  ||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db        8226 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 8285

Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db        8286 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 8345

Qy         121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               |||   | :|:|||:||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db        8346 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 8405


RESULT 6
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-1

  Query Match             61.9%;  Score 546;  DB 3;  Length 4551;
  Best Local Similarity   62.8%;  Pred. No. 2.3e-51;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy           1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
               |:||:||||| ||  ||||: |  ||| |||||||| |||||: |  || ||||:| : |
Db        4038 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4097

Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db        4098 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4157

Qy         121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               |||     :|:|||:||||| : | |||| |||||| |||||:|||||||| :|| ||||
Db        4158 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4217


RESULT 7
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             61.9%;  Score 546;  DB 4;  Length 4613;
  Best Local Similarity   62.8%;  Pred. No. 2.4e-51;
  Matches  113;  Conservative   20;  Mismatches   45;  Indels    2;  Gaps    1;

Qy           1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
               |:||:||||| ||  ||||: |  ||| |||||||| |||||: |  || ||||:| : |
Db        4100 GTVLLTGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAA 4159

Qy          61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db        4160 CDVADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLD 4219

Qy         121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
               |||     :|:|||:||||| : | |||| |||||| |||||:|||||||| :|| ||||
Db        4220 ELTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGLW 4279


RESULT 8
US-08-804-227C-2
; Sequence 2, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
```

```
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4472 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-2

  Query Match             60.5%;  Score 534;  DB 2;  Length 4472;
  Best Local Similarity   64.0%;  Pred. No. 5e-50;
  Matches  114;  Conservative   12;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||| || ||  :|||||   ||| |||:|| |  | ||: | |||| | ||    |
Db       4017 GTVLVTGATGTLGRLLARHLAAEHGVRHLLLLSRGGRAAEGADELAAELAGLEAEPCFAA 4076

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|  : |:|  |||  |||||:|||||||| |: :|||||:|||   | |||| |
Db       4077 CDAADREALARVLAEVPADRPLTGVIHAAGVLDDGTLDALTPERIGTVMRPKADAALNLH 4136

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||   || | :|| ||||||  || ||||||  |||||  ||| ||| ||||||||
Db       4137 ELTRTSPLSVFAVFSGAAGILGRPGQANYAAANTFLDALAQHRRAHGLPAVSLAWGLW 4194


RESULT 9
US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-2

  Query Match             60.1%;  Score 530;  DB 4;  Length 5215;
  Best Local Similarity   62.9%;  Pred. No. 1.7e-49;
  Matches  112;  Conservative   14;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||| :||||    ||| |||||||| |||||  |  :|| |||||   |
Db       2440 GTVLVTGGTGGLGALLARHLVERHGVRRLLLVSRRGADAPGAADLGEDLAGLGAEVAFAA 2499

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
               |  :|  : | :  |||  ||| ||| |||:|||| :||||||  |   |||||  |
Db       2500 ADAADRESLARAIATVPAEHPLTAVVHTAGVVDDATVEALTPERLDAVLRPKVDAAWNLH 2559

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||: : | ||||||| :|||:|:||| ||||||  |||||  | | ||   |||||||
Db       2560 ELTKDLRLDAFVLFSSVSGIVGTAGQANYAAANTGLDALAAHRAATGLAATSLAWGLW 2617


RESULT 10
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match             59.0%;  Score 520;  DB 4;  Length 5588;
  Best Local Similarity   62.9%;  Pred. No. 2.5e-48;
  Matches  112;  Conservative   16;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||| :||||    ||| ||| ||||  ||||: | |||   ||:| :||
Db       4933 GTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGADELVAELEQAGADVAVVA 4992

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|  : ||:  |||  ||  ||||||||||  : |::||||  |   |||||  |
Db       4993 CDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSPERLDAVLRPKVDAAWYLH 5052
```

```
Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| : ||||||||| ||: | ||| |||| || |||||: |:| ||| :||| |||
Db       5053 ELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHCRQAQGLPALSLASGLW 5110


RESULT 11
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match             59.0%;  Score 520;  DB 4;  Length 5588;
  Best Local Similarity   62.9%;  Pred. No. 2.5e-48;
  Matches  112;  Conservative   16;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||| :||||     ||| ||| |||| ||||: | |||    ||:| :||
Db       4933 GTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGADELVAELEQAGADVAVVA 4992

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|  : ||:  |||  ||  ||||||||||  : |::||||  |   |||||  |
Db       4993 CDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSPERLDAVLRPKVDAAWYLH 5052

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| : ||||||||| ||: | ||| |||| || |||||: |:| ||| :||| |||
Db       5053 ELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHCRQAQGLPALSLASGLW 5110


RESULT 12
US-08-804-227C-3
; Sequence 3, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1864 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-3

  Query Match             58.7%;  Score 518;  DB 2;  Length 1864;
  Best Local Similarity   60.7%;  Pred. No. 8.7e-49;
  Matches  108;  Conservative   16;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| |||  ||||     | | ||| || || ||||  | ||||||||   : |
Db       1414 GTVLITGGTGSLGAEAARHLVTRHGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAA 1473

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  ||: |||| ||| |:| ||||||  : : ||:||  |   | ||   |
Db       1474 CDVADRAALRALLDRVPAGHPLTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLH 1533

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||:|  ||||:|:|||||:||||||  |||||| ||: |  ||: ||| ||| || |
Db       1534 ELTQGHALSAFILYSSAAGVLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPW 1591


RESULT 13
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 6095
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-144-085-2

  Query Match             58.1%;  Score 512.5;  DB 4;  Length 6095;
  Best Local Similarity   60.9%;  Pred. No. 2e-47;
  Matches  109;  Conservative   16;  Mismatches   53;  Indels    1;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAE-VRIV 59
              |:||||||| || ||| ||    ||| |:| |||| :|||| |    |  |||| | :
Db       4756 GTVLVTGGTGELGQAVAEHLVRAHGVRHLVLTSRRGLEAPGAPGFVQALEKLGAETVTVA 4815

Qy         60 ACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
              |||| :| || |:| |: |  ||| |:| |||||  | : |||||  | | ||: || |
Db       4816 ACDVSKREEVARVLAGIEAAHPLTAVLHLAGVLDDGVITAQTPERLSRVLAPKVNGALHL 4875

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||  ::|||||||| :| ||:||| ||||||: ||| |  ||: ||   |||| |
Db       4876 HELTEDLDLSAFVLFSSMSGTLGTAGQSNYAAANSFLDAFAAHRRSRGLAATSLAWGFW 4934
```

```
RESULT 14
US-09-144-085-1
; Sequence 1, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 5087
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-144-085-1

  Query Match             57.8%;  Score 509.5;  DB 4;  Length 5087;
  Best Local Similarity   61.5%;  Pred. No. 3.3e-47;
  Matches  110;  Conservative   13;  Mismatches   55;  Indels    1;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAE-VRIV 59
              |:||||||| || ||| ||    ||| |:| |||| :|||| |   || |||| | :
Db       4642 GTVLVTGGTGELGQAVAAHLVRAHGVRHLVLTSRRGLEAPGARELVESLAELGAETVTVA 4701

Qy         60 ACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 119
              |||| :| || |:| |: |  ||: |:| || |||  :|  | |||  | | ||| || |
Db       4702 ACDVSKREEVARVLAGIDAARPLSAVLHLAGALDDGVLAGQTAERLSRVLAPKVDGALHL 4761

Qy        120 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::| ||||||| ||  |:||| ||||||  |||||  ||  ||   |||||||
Db       4762 HELTRELDLVAFVLFSSVAGTFGTAGQSNYAAANTFLDALAAHRRGCGLAATSLAWGLW 4820


RESULT 15
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4928 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-5

  Query Match             57.1%;  Score 504;  DB 4;  Length 4928;
  Best Local Similarity   59.0%;  Pred. No. 1.3e-46;
  Matches  105;  Conservative   19;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||||    |:| |:|  ||| :|||   |  |||  || | :||
Db       1421 GTVLITGGTGALGGVVARHLVEEHGIRRLVLAGRRGWNAPGVHELVDELARAGAVVEVVA 1480

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :| ::  :|  :|   || |:|| |||| |  | ||:   :||||| ||  |  |
Db       1481 CDVADRTDLEHVLAAIPVDWPLRGIVHTAGVLADGVIGSLSAADVGTVFAPKVTGAWHLH 1540

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::|| |||||| :|| |:||| ||||||  |||||  ||| |||||:|||||||
Db       1541 ELTRDLDLSPFVLFSSFSGIAGAAGQANYAAANTFLDALARYRRARGLPGLSLAWGLW 1598


Search completed: June 17, 2003, 13:09:31
Job time : 4.23524 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 8.71975 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_3537_3714
Perfect score: 882
Sequence: 1 GSVLVTGGTGVLGAAVARHL.....LAYRRRAAGLPGVSLAWGLW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	882	100.0	3972	21	AAB23749	S. avermitilis ave
2	882	100.0	3972	22	AAG65264	Streptomyces averm
3	882	100.0	3972	22	AAG65268	Streptomyces averm
4	882	100.0	5532	21	AAB23752	S. avermitilis ave
5	882	100.0	5532	22	AAG65267	Streptomyces averm
6	616	69.8	6239	21	AAB23750	S. avermitilis ave
7	616	69.8	6239	22	AAG65265	Streptomyces averm
8	568	64.4	3724	18	AAW23718	Platenolide syntha
9	568	64.4	3724	18	AAW22608	Platenolide syntha
10	554	62.8	2066	22	AAE10146	Streptomyces nours
11	549	62.2	11096	22	AAE10129	Streptomyces nours
12	548.5	62.2	9477	22	AAE10144	Streptomyces nours
13	547	62.0	3739	21	AAB18638	Amino acid sequenc
14	547	62.0	3739	21	AAY77193	S. venezuelae macr
15	547	62.0	3739	21	AAY77201	S. venezuelae pik
16	547	62.0	3739	21	AAY67202	Narbonolide syntha
17	547	62.0	12199	21	AAY77180	S. venezuelae pik
18	546	61.9	4551	21	AAB18637	Amino acid sequenc
19	546	61.9	4551	21	AAY67201	Narbonolide syntha
20	546	61.9	4613	21	AAY77192	S. venezuelae macr
21	546	61.9	4613	21	AAY77200	S. venezuelae pik
22	546	61.9	5435	22	AAE10145	Streptomyces nours
23	537	60.9	899	22	AAE10153	Protein encoded by
24	534	60.5	4472	18	AAW22601	Tylactone synthase
25	530	60.1	4630	21	AAY77177	S. venezuelae vep
26	523	59.3	4630	18	AAW19629	Streptomyces venez
27	520	59.0	4572	19	AAW52845	A. mediterranei ri
28	520	59.0	5588	20	AAY39301	SpnE a polyketide
29	520	59.0	5588	22	AAB70969	S. spinosa protein
30	518	58.7	1864	18	AAW22602	Tylactone synthase
31	517	58.6	7068	22	AAE10142	Streptomyces nours
32	512.5	58.1	6095	22	AAU10701	Amino acid sequenc
33	511	57.9	6797	22	AAB31558	Pimaricin biosynth
34	509.5	57.8	5087	22	AAU10700	Amino acid sequenc
35	504	57.1	4924	22	AAB70968	S. spinosa protein
36	504	57.1	4928	20	AAY39300	SpnD a polyketide
37	503.5	57.1	1891	19	AAW52844	Amycolatopsis medi
38	503.5	57.1	3413	19	AAW52849	A. mediterranei ri
39	500	56.7	4545	18	AAW22611	Hybrid srmG/tylG O
40	500	56.7	4550	18	AAW23716	Platenolide syntha
41	500	56.7	4550	18	AAW22606	Platenolide syntha
42	495.5	56.2	1688	19	AAW52848	A. mediterranei ri
43	492	55.8	5069	19	AAW52846	A. mediterranei ri
44	491	55.7	4881	21	AAB23751	S. avermitilis ave
45	491	55.7	4881	22	AAG65266	Streptomyces averm

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 882; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1.8e-76;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 3537 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 3596

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 3597 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 3656

Qy 121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 3657 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 3714

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 882; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1.8e-76;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
Db 3537 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 3596

Qy 61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
Db 3597 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 3656

Qy 121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
Db 3657 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 3714

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.

```
XX
SQ   Sequence   3972 AA;

  Query Match           100.0%;  Score 882;  DB 22;  Length 3972;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-76;
  Matches  178;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3537 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 3596

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3597 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 3656

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3657 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 3714


RESULT 4
AAB23752
ID   AAB23752 standard; Protein; 5532 AA.
XX
AC   AAB23752;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 278-305; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;

  Query Match           100.0%;  Score 882;  DB 21;  Length 5532;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-76;
  Matches  178;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1488 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 1547

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1548 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 1607

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1608 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 1665


RESULT 5
AAG65267
ID   AAG65267 standard; Protein; 5532 AA.
XX
AC   AAG65267;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 7.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 217-235; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   5532 AA;

  Query Match           100.0%;  Score 882;  DB 22;  Length 5532;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-76;
  Matches  178;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1488 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 1547

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1548 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 1607

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1608 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 1665


RESULT 6
AAB23750
ID   AAB23750 standard; Protein; 6239 AA.
XX
AC   AAB23750;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92301.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 223-254; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   6239 AA;

  Query Match             69.8%;  Score 616;  DB 21;  Length 6239;
  Best Local Similarity   69.1%;  Pred. No. 2.4e-50;
  Matches  123;  Conservative   16;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||||||  |||||    ||| |||  |||||| ||  |||||  ||| | :||
Db       5753 GTVLITGGTGVLGRLVARHLVEAHGVRHLLLAGRRGPDAEGAPELRAELGGLGATVEVVA 5812

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|:::  ||  :|   |||||||:||:|||  | ||:|||||  |  || |||||||
Db       5813 CDAADRQQLADLLTRIPDDRPLTGVVHSAGILDDGVITSLSPERLGAVLRAKADAALLLD 5872

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||| ||||||:||||: ::|| |||||||||| || ||:|||| ||| ||||||||
Db       5873 ELTRGAELSAFVMFSSASAVVGSPGQGNYAAANAVLDFLAHRRRAEGLPAVSLAWGLW 5930


RESULT 7
AAG65265
ID   AAG65265 standard; Protein; 6239 AA.
XX
AC   AAG65265;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 5.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79277.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 180-201; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   6239 AA;

  Query Match             69.8%;  Score 616;  DB 22;  Length 6239;
  Best Local Similarity   69.1%;  Pred. No. 2.4e-50;
  Matches  123;  Conservative   16;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||||||  |||||    ||| |||  |||||| ||  |||||  ||| | :||
Db       5753 GTVLITGGTGVLGRLVARHLVEAHGVRHLLLAGRRGPDAEGAPELRAELGGLGATVEVVA 5812

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||  :|:::  ||  :|   |||||||:||:|||  | ||:|||||  |  || |||||||
Db       5813 CDAADRQQLADLLTRIPDDRPLTGVVHSAGILDDGVITSLSPERLGAVLRAKADAALLLD 5872

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||| ||||||:||||: ::|| |||||||||| || ||:|||| ||| ||||||||
Db       5873 ELTRGAELSAFVMFSSASAVVGSPGQGNYAAANAVLDFLAHRRRAEGLPAVSLAWGLW 5930


RESULT 8
AAW23718
ID   AAW23718 standard; Protein; 3724 AA.
XX
AC   AAW23718;
XX
```

```
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase ORF3 protein.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "encoded by TGT"
FT   Domain          36..459
FT                   /note= "ketosynthase domain, KS4"
FT   Domain          594..921
FT                   /note= "acyltransferase domain, AT4"
FT   Domain          1177..1350
FT                   /note= "ketosynthase domain, KS04"
FT   Domain          1459..1523
FT                   /note= "acyl carrier protein domain, ACP4"
FT   Domain          1545..1969
FT                   /note= "ketosynthase domain, KS5"
FT   Domain          2088..2424
FT                   /note= "acyltransferase domain, AT5"
FT   Domain          2445..2619
FT                   /note= "dehydratase domain, DH5"
FT   Domain          2912..3261
FT                   /note= "enoylreductase domain, ER5"
FT   Domain          3266..3451
FT                   /note= "ketoreductase domain, KR5"
FT   Domain          3551..3631
FT                   /note= "acyl carrier protein domain, ACP5"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   N-PSDB; AAT78508.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 8; Pages 54-66; 81pp; English.
XX
CC   AAW23716-W23720 represent proteins encoded by the platenolide synthase
CC   gene cluster of the invention. The gene cluster is also referred to as
CC   the srmG gene, and was isolated from Streptomyces ambofaciens. These
CC   sequences are multi-functional proteins which direct the synthesis of
CC   the polyketide platenolide. Platenolide is the basic building block of
CC   the macrolide antibiotic spiramycin. The DNA can be used to produce
CC   compounds exhibiting antibiotic activity based on the platenolide
CC   structure, including specifically the macrolide antibiotic spiramycin and
CC   spiramycin analogues and derivatives. Modifications of the platenolide
CC   synthase DNA sequence can be made so as to change the number and type of
CC   carboxylic acids incorporated into the growing polyketide chain and to
CC   change the kind of post-condensation processing that is conducted.
XX
SQ   Sequence   3724 AA;

  Query Match           64.4%;  Score 568;  DB 18;  Length 3724;
  Best Local Similarity  65.2%;  Pred. No. 6.2e-46;
  Matches  116;  Conservative   19;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||: :|||||   ||| ||: :|||||  ||  | |:||||||   :||
Db       3268 GTVLVTGGTGALGSRIARHLASRHGVRRLLIAARRGPDGEGAAELVADLAALGASATVVA 3327

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :   |  || |:|| ||| |||: ||||| : |||||:  |   ||:||: ||
Db       3328 CDVSDADAVRGLLAGIPADHPLTAVVHSTGVLDDGVLPGLTPERMRRVLRPKVEAAVHLD 3387

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::|||||||||:||:|||  ||||||||| ||||| |||: ||| ||||||||
Db       3388 ELTRDLDLSAFVLFSSSAGLLGSPAQGNYAAANATLDALAARRRSLGLPSVSLAWGLW 3445


RESULT 9
AAW22608
ID   AAW22608 standard; Protein; 3724 AA.
XX
AC   AAW22608;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase ORF3 protein.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "encoded by TGT"
FT   Domain          36..459
FT                   /note= "ketosynthase domain, KS4"
FT   Domain          594..921
FT                   /note= "acyltransferase domain, AT4"
FT   Domain          1177..1350
FT                   /note= "ketosynthase domain, KS04"
FT   Domain          1459..1523
FT                   /note= "acyl carrier protein domain, ACP4"
FT   Domain          1545..1969
FT                   /note= "ketosynthase domain, KS5"
FT   Domain          2088..2424
FT                   /note= "acyltransferase domain, AT5"
FT   Domain          2445..2619
FT                   /note= "dehydratase domain, DH5"
FT   Domain          2912..3261
FT                   /note= "enoylreductase domain, ER5"
FT   Domain          3266..3451
FT                   /note= "ketoreductase domain, KR5"
FT   Domain          3551..3631
FT                   /note= "acyl carrier protein domain, ACP5"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80414.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 155-167; 220pp; English.
XX
CC   AAW22606-W22610 represent proteins encoded by the platenolide synthase
```

```
CC   gene cluster. The gene cluster is also referred to as the srmG gene, and
CC   was isolated from Streptomyces ambofaciens. These sequences are
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA encoding this sequence was used along
CC   with the tylG gene (see AAT80413) to create a hybrid ORF1 sequence (see
CC   AAT80415). The tylG gene is the tylactone synthase gene cluster of the
CC   invention. The tylG sequence was isolated from Streptomyces fradiae, and
CC   encodes multifunctional proteins which direct the synthesis of the
CC   polyketide tylactone. Tylactone is the basic building block of the
CC   antibiotic tylosin. The hybrid sequence can be used to transform S.
CC   ambofaciens lacking the srmG ORF1 sequence, or S. fradiae lacking the
CC   tylG ORF1 sequence, so that they can produce polyketides. The DNA
CC   sequence can be modified so as to alter the type of carboxylic acids
CC   incorporated, the number of carboxylic acids incorporated and/or the
CC   post-condensation reactions performed, thereby resulting in novel
CC   tylosin-related polyketides.
XX
SQ   Sequence   3724 AA;

  Query Match             64.4%;  Score 568;  DB 18;  Length 3724;
  Best Local Similarity   65.2%;  Pred. No. 6.2e-46;
  Matches  116;  Conservative   19;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||: :|||||   ||| ||: :|||||  ||  | |:||||||   :||
Db       3268 GTVLVTGGTGALGSRIARHLASRHGVRRLLIAARRGPDGEGAAELVADLAALGASATVVA 3327

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :   |  || |:||  ||| |||: |||||  :  |||||:  |   ||:||: ||
Db       3328 CDVSDADAVRGLLAGIPADHPLTAVVHSTGVLDDGVLPGLTPERMRRVLRPKVEAAVHLD 3387

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||| ::|||||||||:||:|||  |||||||| ||||| |||: ||| ||||||||
Db       3388 ELTRDLDLSAFVLFSSSAGLLGSPAQGNYAAANATLDALAARRRSLGLPSVSLAWGLW 3445


RESULT 10
AAE10146
ID   AAE10146 standard; Protein; 2066 AA.
XX
AC   AAE10146;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysK.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          34..457
FT                   /label= KS18_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          568..881
FT                   /label= AT18_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          898..1102
FT                   /label= DH18_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1416..1663
FT                   /label= KR18_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1695..1769
FT                   /label= ACP18_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1849..2066
FT                   /label= TE_domain
FT                   /note= "Thioesterase (TE) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 263-264; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin gene,
CC   NysK.
XX
SQ   Sequence   2066 AA;

  Query Match             62.8%;  Score 554;  DB 22;  Length 2066;
  Best Local Similarity   64.0%;  Pred. No. 7.2e-45;
  Matches  114;  Conservative   18;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||||:|||| | | | |||  ||||:||||  | ||| | ||:||: |
Db       1416 GTVLITGGTGGLGAALARHLVTVRGARHLLLAGRRGPEAPGAGELVAELTAQGADVRVAA 1475

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||||:|  :  ||  |||  ||| ||| ||||||| | ||||::| ||   | |||  |
Db       1476 CDVGDRTALDALLATVPAAHPLTAVVHTAGVLDDALIGSLTPDQLATVLRPKADAAWHLH 1535

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              : |||::|: |||:|| :|:||| |||||||||| |||||  |   ||| :||||| |
Db       1536 DATRGLDLAGFVLYSSVSGVLGSPGQGNYAAANAYLDALARHRADQGLPALSLAWGPW 1593


RESULT 11
AAE10129
ID   AAE10129 standard; Protein; 11096 AA.
XX
AC   AAE10129;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysC.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
```

```
FT   Domain          35..455
FT                   /label= KS3_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          546..858
FT                   /label= AT3_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          872..1073
FT                   /label= DH3_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1381..1628
FT                   /label= KR3_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1662..1735
FT                   /label= ACP3_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1757..2180
FT                   /label= KS4_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2291..2603
FT                   /label= AT4_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2617..2818
FT                   /label= DH4_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3124..3371
FT                   /label= KR4_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3407..3480
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3501..3924
FT                   /label= KS5_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4032..4346
FT                   /label= AT5_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4360..4561
FT                   /label= DH5_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4953..5239
FT                   /label= ER5_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          5248..5495
FT                   /label= KR5_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5528..5601
FT                   /label= ACP5_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          5623..6046
FT                   /label= KS6_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          6166..6478
FT                   /label= AT6_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          6492..6704
FT                   /label= DH6_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          7038..7281
FT                   /label= KR6_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          7315..7388
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          7408..7831
FT                   /label= KS7_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          7939..8253
FT                   /label= AT7_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          8267..8470
FT                   /label= DH7_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          8812..9086
FT                   /label= KR7_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          9120..9193
FT                   /label= ACP7_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          9214..9637
FT                   /label= KS8_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          9758..10072
FT                   /label= AT8_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          10086..10289
FT                   /label= DH8_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          10657..10904
FT                   /label= KR8_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          10939..11012
FT                   /label= ACP8_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17184.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 170-176; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin
CC   gene, NysC.
XX
SQ   Sequence   11096 AA;

  Query Match             62.2%;  Score 549;  DB 22;  Length 11096;
  Best Local Similarity   65.2%;  Pred. No. 1.6e-43;
  Matches  116;  Conservative   12;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:|||| ||| |||| |||:|||||||||||||  | ||| | |||| : |
Db       8812 GTVLITGGTGGLGAQFARHLVDRYGVRNLLLVSRRGPDAPGTTELVAELTAHGAEVAVQA 8871

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :   |  |: |||   ||  ||| ||||||  | ||| ||| ||   | |||  |
Db       8872 CDVADGDAVAALVAGVPDEHPLRAVVHTAGVLDDGVIGSLTEERLATVLRPKADAAWHLH 8931
```

```
Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              | |||::| |||:||| ||: | ||| ||||||| |||| :|||||||:||||| |
Db       8932 EATRGLDLDAFVVFSSVAGVFGGAGQANYAAANAFLDALMAQRRAAGLPGLSLAWGPW 8989


RESULT 12
AAE10144
ID   AAE10144 standard; Protein; 9477 AA.
XX
AC   AAE10144;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (complete).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          41..464
FT                   /label= KS15_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          578..889
FT                   /label= AT15_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          903..1102
FT                   /label= DH15_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1446..1731
FT                   /label= ER15_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          1740..1988
FT                   /label= KR15_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          2023..2096
FT                   /label= ACP15_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          2117..2538
FT                   /label= KS16_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2635..2953
FT                   /label= AT16_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2967..3167
FT                   /label= DH16_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3257..3500
FT                   /label= KR16_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3539..3612
FT                   /label= ACP16_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3634..4057
FT                   /label= KS17_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4153..4472
FT                   /label= AT17_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4486..4725
FT                   /label= DH17_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4997..5245
FT                   /label= KR17_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5277..5350
FT                   /label= ACP17_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 255-260; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (complete).
XX
SQ   Sequence   9477 AA;

  Query Match             62.2%;  Score 548.5;  DB 22;  Length 9477;
  Best Local Similarity   63.5%;  Pred. No. 1.5e-43;
  Matches  113;  Conservative   21;  Mismatches   43;  Indels    1;  Gaps    1;

Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||||||| ||| ||| ||   | : |:|:|||||||||  ||||| ||| :| : |
Db       9050 GTVLVTGGTGALGAHVARRLA-KDGAQHLVLLSRRGPDAPGAAELRAELDALGTDVTVAA 9108

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :| ::  :|: :||  ||||||| ||||||  :  |||||   || ||| :|||||
Db       9109 CDVADRDQLTAVLDALPADRPLTGVVHTAGVLDDGVLDRLTPERFQEVFRAKVTSALLLD 9168

Qy        121 ELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              ||||  ||:|||||||:  :|: || ||||||| ||||| :||  |||  |::|| |
Db       9169 ELTRDRELAAFVLFSSASAAVGNPGQANYAAANAVLDALAEQRRVLGLPATSVSWGAW 9226


RESULT 13
AAB18638
ID   AAB18638 standard; Protein; 3739 AA.
XX
AC   AAB18638;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of narbonolide synthase subunit 2 (PICAII).
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis.
XX
OS   Streptomyces venezuelae.
XX
```

PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 11-12; 117pp; English.
XX
CC The present sequence represents a narbonolide synthase subunit 2
CC (PICAII). The nucleotide sequence encoding it is used in the course of
CC the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 3739 AA;

Query Match 62.0%; Score 547; DB 21; Length 3739;
Best Local Similarity 62.8%; Pred. No. 6.9e-44;
Matches 113; Conservative 19; Mismatches 46; Indels 2; Gaps 1;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456
```

RESULT 14
AAY77193
ID AAY77193 standard; Protein; 3739 AA.
XX
AC AAY77193;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae macrolide biosynthetic enzyme PikAII, SEQ ID NO:33.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87298.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 19; Page 403-415; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC AAZ87295-Z87302.
XX
SQ Sequence 3739 AA;

Query Match 62.0%; Score 547; DB 21; Length 3739;
Best Local Similarity 62.8%; Pred. No. 6.9e-44;
Matches 113; Conservative 19; Mismatches 46; Indels 2; Gaps 1;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  |||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456
```

RESULT 15
AAY77201
ID AAY77201 standard; Protein; 3739 AA.
XX
AC AAY77201;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster protein #2.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87318.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77200-Y77203 and AAY80997 represent proteins encoded by the
CC macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
CC ATCC 15439, as given in figure 31.
XX
SQ Sequence 3739 AA;

Query Match 62.0%; Score 547; DB 21; Length 3739;
Best Local Similarity 62.8%; Pred. No. 6.9e-44;

Matches 113; Conservative 19; Mismatches 46; Indels 2; Gaps 1;

```
Qy          1 GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVA 60
              |:||:||||| ||  ||||: |  ||| |||||||| |||||  |  || ||||:| : |
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLD 120
              ||| :|  :  :|: :||  ||| ||| |||| | |: |:| | :  |   ||||| |||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 ELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 178
              |||   | :|:|||:||||| : | |||| |||||| |||||:||| |||| :|| ||||
Db       3397 ELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLW 3456
```

Search completed: June 17, 2003, 13:02:34
Job time : 9.71975 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 14.6048 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_29_344
Perfect score: 1618
Sequence: 1 VFVFPGQGPQWPGMGRELLD.....ADTVIMGTLRRGQGTLDHFL 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1618	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	1578	97.5	3626	2	Q9F7T9	Q9f7t9 streptomyce
3	819	50.6	3654	2	O30766	O30766 streptomyce
4	817.5	50.5	8817	2	Q53840	Q53840 polyangium
5	791	48.9	4881	2	Q9S0R3	Q9s0r3 streptomyce
6	773	47.8	4151	16	O53490	O53490 mycobacteri
7	765.5	47.3	3192	2	Q9L4W4	Q9l4w4 streptomyce
8	765	47.3	2595	2	Q9ALM6	Q9alm6 saccharopol
9	763	47.2	3939	2	Q93HJ3	Q93hj3 streptomyce
10	762	47.1	3729	2	O33956	O33956 streptomyce
11	758	46.8	3970	2	Q93HI8	Q93hi8 streptomyce
12	756.5	46.8	4498	2	Q93HE5	Q93he5 streptomyce
13	756	46.7	9510	2	Q93NX9	Q93nx9 streptomyce
14	754.5	46.6	9507	2	Q9EWA1	Q9ewa1 streptomyce
15	753	46.5	9477	2	Q9L4X3	Q9l4x3 streptomyce
16	750.5	46.4	3190	2	Q93NW7	Q93nw7 streptomyce
17	750	46.4	5588	2	Q9ALM2	Q9alm2 saccharopol
18	746	46.1	3562	2	Q9F829	Q9f829 micromonosp
19	745.5	46.1	1041	2	Q93C90	Q93c90 actinomadur
20	740	45.7	3352	2	Q93H83	Q93h83 streptomyce
21	737	45.6	4685	2	Q93HJ2	Q93hj2 streptomyce
22	734	45.4	5532	2	Q9S0R4	Q9s0r4 streptomyce
23	733	45.3	3613	2	Q93HJ1	Q93hj1 streptomyce
24	731	45.2	6146	2	Q93HJ5	Q93hj5 streptomyce
25	722.5	44.7	6315	2	Q9ADL6	Q9adl6 polyangium
26	721	44.6	1907	2	Q93HI9	Q93hi9 streptomyce
27	718.5	44.4	3201	2	Q9F828	Q9f828 micromonosp
28	714.5	44.2	3816	2	Q9KIV3	Q9kiv3 streptomyce
29	713	44.1	6239	2	Q9S0R7	Q9s0r7 streptomyce
30	712	44.0	4150	2	Q9KIV4	Q9kiv4 streptomyce
31	706	43.6	4840	2	Q93HJ4	Q93hj4 streptomyce
32	704	43.5	3546	2	Q9F830	Q9f830 micromonosp
33	699	43.2	1937	2	O30482	O30482 streptomyce
34	697.5	43.1	5060	2	O52545	O52545 amycolatops
35	697.5	43.1	5069	2	O52789	O52789 amycolatops
36	696.5	43.0	4613	2	Q9ZGI5	Q9zgi5 streptomyce
37	689.5	42.6	2126	16	P94996	P94996 mycobacteri
38	685	42.3	3524	2	Q93H86	Q93h86 streptomyce
39	681.5	42.1	3739	2	Q9ZGI4	Q9zgi4 streptomyce
40	674.5	41.7	3170	2	Q9ALM4	Q9alm4 saccharopol
41	674	41.7	1562	2	Q9ZGI3	Q9zgi3 streptomyce
42	673.5	41.6	1762	2	O52546	O52546 amycolatops
43	673.5	41.6	1763	2	O52790	O52790 amycolatops
44	672	41.5	7525	2	Q9KIE0	Q9kie0 streptomyce
45	667.5	41.3	1346	2	Q9ZGI2	Q9zgi2 streptomyce

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 1618; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 3.5e-100;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         29 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 88

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 148

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 208

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 268

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 328

Qy        301 IMGTLRRGQGTLDHFL 316
              ||||||||||||||||
Db        329 IMGTLRRGQGTLDHFL 344
```

RESULT 2
Q9F7T9
ID Q9F7T9 PRELIMINARY; PRT; 3626 AA.
AC Q9F7T9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Avermectin polyketide synthase (Fragment).
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31271;
RA Hong Y.-S., Lee J.J.;
RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT and polyketide synthase gene from Streptomyces avermitilis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275943; AAG09812.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT NON_TER 3626 3626
SQ SEQUENCE 3626 AA; 380557 MW; 6272F5F088C1A8D0 CRC64;

Query Match 97.5%; Score 1578; DB 2; Length 3626;
Best Local Similarity 98.1%; Pred. No. 1.6e-97;
Matches 310; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         10 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAEFAPYVDWSVEQVLRDSPDAPGLDRV 69

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||   |||||
Db         70 DVVQRTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVGDAWSQA 129

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        130 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 189

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 249

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAADLGDTTGTADTV 309

Qy        301 IMGTLRRGQGTLDHFL 316
              ||||||||||||||||
Db        310 IMGTLRRGQGTLDHFL 325
```

RESULT 3
O30766
ID O30766 PRELIMINARY; PRT; 3654 AA.
AC O30766;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyketide synthase modules 4 and 5.
GN NIDA3.
OS Streptomyces caelestis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=36816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL-2821;
RX MEDLINE=98053867; PubMed=9393718;
RA Kakavas S.J., Katz L., Stassi D.;
RT "Identification and characterization of the niddamycin polyketide
RT synthase genes from Streptomyces caelestis.";
RL J. Bacteriol. 179:7515-7522(1997).
DR EMBL; AF016585; AAC46026.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR001993; Mitoch_carrier.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR002364; QOR_zeta_crystal.
DR Pfam; PF00698; Acyl_transf; 2.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00032; ANTENNAPEDIA; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3654 AA; 379108 MW; 59825E49EDCD06FF CRC64;

Query Match 50.6%; Score 819; DB 2; Length 3654;
Best Local Similarity 54.0%; Pred. No. 2.4e-46;
Matches 168; Conservative 38; Mismatches 95; Indels 10; Gaps 3;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||   ||: |  || :  :|:||  ||: |||  |||  |||| ||||
Db       2008 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 2067

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| || :|:||||:||: |||| ||:||| |||||||:| ||| |:||:| | |:|
Db       2068 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 2127

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | :|:|    :|| ||:  | :    |||||||| | |  |:|  ||:|:|||
Db       2128 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 2183
```

```
Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              ||    |  | || ||   ||| : |:   ::  | |: |||: |||:|:|||| ||  |
Db       2184 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTE 2243

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||| ||||||   | || | | |:  |   |:| ||||:| : |::  | |      |
Db       2244 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 2298

Qy        300 VIMGTLRRGQG 310
              ::|||||  |
Db       2299 AVLGTLRRRHG 2309


RESULT 4
Q53840
ID   Q53840      PRELIMINARY;      PRT;  8817 AA.
AC   Q53840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Soraphen polyketide synthase B.
GN   SORB.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE OF 6342-8466 FROM N.A.
RC   STRAIN=SO CE26;
RX   MEDLINE=95325306; PubMed=7601830;
RA   Schupp T., Toupet C., Cluzel B., Neff S., Hill S., Beck J.J.,
RA   Ligon J.M.;
RT   "A Sorangium cellulosum (myxobacterium) gene cluster for the
RT   biosynthesis of the macrolide antibiotic soraphen A: cloning,
RT   characterization, and homology to polyketide synthase genes from
RT   actinomycetes.";
RL   J. Bacteriol. 177:3673-3679(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SO CE26;
RA   Ligon J.M., Hill S., Beck J., Zirkle R., Molnar I., Zawodny J.,
RA   Money S., Schupp T.;
RT   "Characterization of the biosynthetic gene cluster for the antifungal
RT   polyketide soraphen A from Sorangium cellulosum So ce26.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U24241; AAA79984.2; -.
DR   HSSP; P24249; 1HNK.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR003662; sub_transporter.
DR   InterPro; IPR001031; Thioesterase.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00698; Acyl_transf; 5.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 5.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF00975; Thioesterase; 1.
DR   TIGRFAMs; TIGR00128; fabD; 5.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 5.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_3.
KW   Phosphopantetheine.
SQ   SEQUENCE   8817 AA;  933938 MW;  125E1F51F35CDBC1 CRC64;

  Query Match           50.5%;  Score 817.5;  DB 2;  Length 8817;
  Best Local Similarity   53.5%;  Pred. No. 9.4e-46;
  Matches  169;  Conservative   42;  Mismatches   96;  Indels    9;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||| || ||   |||:| |||  : ||| | ||:|:||:  |||    || ||||
Db       7311 VFVFPGQGSQWEGMALSLLDSSPVFRAQLEACERALAPHVEWSLLAVLRRDEGAPSLDRV 7370

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:|||||||| |||| ||:||| |||||| |:| ||| ||||:  | |:|
Db       7371 DVVQPALFAVMVSLAALWRSLGVEPAAVVGHSQGEIAAAFVAGALSLEDAARIAALRSKA 7430

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||:|| ||: :|     :|   :||| :    ||::|||| ||:|:|||   |:  |:
Db       7431 LTTVAGNGAMAAVELGASDLQTYLAPWGD----RLSIAAVNSPRATLVSGEPAAIDALID 7486

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              ||||||  | : ||  :||  | |::: : :||  | ||   :| :|:||| |||  ||
Db       7487 SLTAAQVFARKVRVDYASHSAQMDAVQDELAAGLANIAPRTCELPLYSTVTGTRLDGSEL 7546

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | ||||||:  || |  |   ||  | : |||:|||||||: |:|          | |
Db       7547 DGAYWYRNLRQTVLFSSATERLLDDGHRFFVEVSPHPVLTLALRETC-----ERSPLDPV 7601

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|::|| :| |   |
Db       7602 VVGSIRRDEGHLARLL 7617


RESULT 5
Q9S0R3
ID   Q9S0R3      PRELIMINARY;      PRT;  4881 AA.
AC   Q9S0R3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 4.
GN   AVEA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84479.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4881 AA;  510303 MW;  F6568C1F01ACD37A CRC64;

  Query Match           48.9%;  Score 791;  DB 2;  Length 4881;
  Best Local Similarity   52.2%;  Pred. No. 2.7e-44;
  Matches  164;  Conservative   45;  Mismatches   97;  Indels    8;  Gaps    2;

Qy          3 VFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDV 62
              :| |||  | || :||| :| || : :  |  |  ||:|||:  ||| :|||| | ||||
Db       2139 LFGGQGTHWEGMAQELLGSSPVFAQQMSDCAQALEPYLDWSLLDVLRGAPDAPPLQRVDV 2198
```

```
Qy          63 VQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQT 122
               ||| |||||:|||||||| || | || ||| |||||||:|:| ||| |||||  | |||
Db        2199 VQPVLFAVMVSLAALWRSYGVHPDAVAGHSQGEIAAAYVAGALSLDDAARVTALRSQALA 2258

Qy         123 TLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADL 182
               ||| ||: ||    ::| ||:| | :    || :||||| || |||| ||| || :|
Db        2259 ALAGQGAMASVGLPVEKLEPRLATWGD----RLVIAAVNGARSAVVSGEPEAVDALVEEL 2314

Qy         183 TAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDA 242
               :   |  | : ||  :||| : ||: |::  | ||  |   :||:|:||| |:|  ||||
Db        2315 SHEDVPARRLMVDWASHSPQVEAIQGRLLELLAPIRARTGDVPFYSTVTGERIDGTELDA 2374

Qy         243 AYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIM 302
               |||||:   |||  | : |:: |   |:|  ||| : :|:||   ::|| |:::: 
Db        2375 DYWYRNLRQVVRFRDATQALVRAGHTVFIEACPHPAVAVGVQETLDEMGD----LDSLVV 2430

Qy         303 GTLRRGQGTLDHFL 316
               |:||||:| |  ||
Db        2431 GSLRRGEGGLRRFL 2444
```

```
RESULT 6
O53490
ID   O53490      PRELIMINARY;      PRT;  4151 AA.
AC   O53490;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase.
GN   PKS12 OR RV2048C OR MT2108 OR MTV018.35C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021899; CAA17262.1; -.
DR   EMBL; AE007061; AAK46387.1; -.
DR   HSSP; P25715; 1MLA.
DR   TIGR; MT2108; -.
DR   TubercuList; Rv2048c; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine; Transferase; Complete proteome.
FT   CONFLICT   2147   2147       H -> Q (IN REF. 2).
FT   CONFLICT   2260   2261       DI -> TV (IN REF. 2).
FT   CONFLICT   2268   2268       W -> H (IN REF. 2).
FT   CONFLICT   2272   2272       S -> A (IN REF. 2).
FT   CONFLICT   2279   2280       AF -> PY (IN REF. 2).
FT   CONFLICT   2282   2283       AA -> GR (IN REF. 2).
FT   CONFLICT   2287   2287       T -> V (IN REF. 2).
FT   CONFLICT   2289   2289       F -> W (IN REF. 2).
FT   CONFLICT   3004   3004       S -> L (IN REF. 2).
FT   CONFLICT   3649   3649       P -> A (IN REF. 2).
SQ   SEQUENCE   4151 AA;  431577 MW;  536F644F111D7A5D CRC64;

  Query Match             47.8%;  Score 773;  DB 16;  Length 4151;
  Best Local Similarity   51.6%;  Pred. No. 3.6e-43;
  Matches  163;  Conservative   45;  Mismatches  102;  Indels    6;  Gaps    2;

Qy           1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
               |||||||| || ||| |||| :  | : : ||  ||| :||||:  ||| :| |||||||
Db         559 VFVFPGQGSQWLGMGIELLDTAPAFAQQIDACAEAFAEFVDWSLVDVLRGAPGAPGLDRV 618

Qy          61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
               ||||| |||||:||| ||:|  | | ||:||| |||||||:|:| ||| |||||||| |:
Db         619 DVVQPVLFAVMVSLAELWKSVAVHPDAVIGHSQGEIAAAYVAGALSLRDAARVVTLRSKL 678

Qy         121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                  ||| | :||:|   |:    :||: :    |:::| |||| : ||||   |: :|:|
Db         679 LAGLAGPGGMVSIACGADQARDLLAPFGD----RVSIAVVNGPSAVVVSGEVGALEELIA 734

Qy         181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
                 :  ::||| | ||  :||  : ||   :   |  | || :|  | |:||| |||| |
Db         735 VCSTKELRTRRIEVDYASHSVEVEAIRGPLAEALSGIEPRSTRTVFFSTVTGNRLDTAGL 794

Qy         241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
               || |||||:  || |: | |   :|| :||:| |||| |  |::|       | | :: :
Db         795 DADYWYRNVRQTVLFDQAVRNACEQGYRTFIESSPHPALITGVEETF--AACTDGDSEAI 852

Qy         301 IMGTLRRGQGTLDHFL 316
               :: || || | |  ||
Db         853 VVPTLGRGDGGLHRFL 868
```

```
RESULT 7
Q9L4W4
ID   Q9L4W4      PRELIMINARY;      PRT;  3192 AA.
AC   Q9L4W4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysB.
GN   NYSB.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71775.1; -.
```

```
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3192 AA;  333916 MW;  F88F8864F086DA64 CRC64;

  Query Match           47.3%;  Score 765.5;  DB 2;  Length 3192;
  Best Local Similarity   50.6%;  Pred. No. 8.3e-43;
  Matches  160;  Conservative   42;  Mismatches  109;  Indels    5;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAPAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || |||||||  |:||| ::  |  | |||||||:  ||     |   |:||
Db       2114 VFVFPGQGSQWWGMGRELLATSEVFRTAIDDCATALAPYVDWSLHDVLAGEGDPALLERV 2173

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||:|: |:||||| || | ||:||| |||||| |:| ||||||||| | |||
Db       2174 DVVQPALFAMMVGLSALWRSHGVVPAAVVGHSQGEIAAACVAGALSLADAARVVALRSQA 2233

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                |:| | ::||:|  : :   :||| |     |:||||||| | ||||  :|:  |
Db       2234 LPQLSGRGGMMSVSAPVERVTALLAPWQE----ALSVAAVNGPSSVVVSGDTDALDALHT 2289

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
                  || | : ||  :|   : |: : :   | |: ||   :||:|:||| |:|
Db       2290 ACQEQGVRARKVSVDYASHGRHVEAVRDELARVLAPVDPRAPEVPFYSTVTGDRVDDAAF 2349

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | |||| |:  ||| | | | ||  | : |:|:|||||   :||    : : ||    |
Db       2350 DGAYWYTNLRQTVRMEEATRALLAAGHRVFIEVSPHPVLAAPIQETQEAVAEATG-GSAV 2408

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|:||| :|    ||
Db       2409 VLGSLRRDEGGPRRFL 2424


RESULT 8
Q9ALM6
ID   Q9ALM6      PRELIMINARY;      PRT;  2595 AA.
AC   Q9ALM6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase loading and extender module 1.
GN   SPNA.
OS   Saccharopolyspora spinosa.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23264.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2595 AA;  273663 MW;  52EF0BBA4D14FA62 CRC64;

  Query Match           47.3%;  Score 765;  DB 2;  Length 2595;
  Best Local Similarity   50.6%;  Pred. No. 6.9e-43;
  Matches  157;  Conservative   51;  Mismatches   92;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAPAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||||:||: |:||  |:  |  | ||::|||:  |||::  |  ||
Db       527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||:|:||| |||| || | ||:||| |||||| |:| ||: |||||| : |:
Db       587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                | |||:||: |:    :|:   : || :    |: || |||| ||:||| |||:|:|:|
Db       647 LTALAGSGAMASLQHPAEEVRQILLPWRD----RIGVAGVNGPSSTLVSGDREAMAELLA 702

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :   ::| | |||:  :||| :  : : ::  | |: ||   || :|: ||  || | :
Db       703 ECADRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPM 761

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:   | || |   ||::|   |:|:|||||  :|| |   |      : |
Db       762 DADYWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAG-----REVV 816

Qy        301 IMGTLRRGQG 310
              :||||||:|
Db       817 ALGTLRRGEG 826


RESULT 9
Q93HJ3
ID   Q93HJ3      PRELIMINARY;      PRT;  3939 AA.
AC   Q93HJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA3.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69194.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
```

```
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3939 AA;  410482 MW;  534AE3B517256503 CRC64;

  Query Match             47.2%;  Score 763;  DB 2;  Length 3939;
  Best Local Similarity   50.2%;  Pred. No. 1.6e-42;
  Matches  157;  Conservative   43;  Mismatches  103;  Indels   10;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || || ||||| | ||      || ||   :||| |  |||: | |  ::|:
Db        572 VFVFPGQGSQWAGMARELLDTSPVFAAKAAACAAAIEEFVDWKVLDVLREKPGAASMERI 631

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQ- 119
              :|:|| || ||:||| |||| |||| ||:| | |||||||:|||:| |||||: | |
Db        632 EVIQPVLFTVMVSLAELWRSYGVEPDAVVGSSQGEIAAAHVAGGLTLQDAARVICLRSAL 691

Qy        120 AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
                || | ||: |||   ||:  |:  | :    ||:|| |||||   |:|  :|:|: :
Db        692 LAETLVGKGAVASVALPADEVRERLTRWAD----RLSVAGVNGPRIVAVAGDDDALAEFI 747

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              |:     || |::   || |  |:  : ||::  | |: ||   :| :|:|||||||
Db        748 AECARDNVRARIVAATVPTHCALVDPLRERLLDLLAPVAPRTGTVPLYSTVTGGRLDTAA 807

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              :|| ||| |  : | |||  | || :|   ||| | |||| ||:::     |  | |
Db        808 MDAGYWYDNTRAPVLFEPVVRALLAEGHSAFVESSAHPVLAMGVEQTV----DAAG-ASG 862

Qy        300 VIMGTLRRGQGTL 312
              |:: :||| :| |
Db        863 VVVESLRRDEGGL 875


RESULT 10
O33956
ID   O33956      PRELIMINARY;      PRT;  3729 AA.
AC   O33956;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Tylactone synthase modules 4 & 5.
GN   TYLG.
OS   Streptomyces fradiae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1906;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DeHoff B.S., Sutton K.L., Rosteck P.R. Jr.;
RT   "Sequence of Streptomyces fradiae tylactone synthase gene tylG.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U78289; AAB66506.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3729 AA;  389375 MW;  19483A58E027BAC1 CRC64;

  Query Match             47.1%;  Score 762;  DB 2;  Length 3729;
  Best Local Similarity   50.9%;  Pred. No. 1.7e-42;
  Matches  161;  Conservative   46;  Mismatches   99;  Indels   10;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| ||| | |:||| :  |||: :||:|| : ::|||| ||:: |||| | ||
Db       2096 VFVFPGQGSQWPSMARDLLDRAPAFRETAKACDAALSVHLDWSVLDVLQEKPDAPPLSRV 2155

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| || :|:|||| ||  || | ||:||| |||||| |:| ||| ||||:| | |:|
Db       2156 DVVQPVLFTMMLSLAACWRDLGVHPAAVVGHSQGEIAAACVAGALSLEDAARIVALRSRA 2215

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              |||| | : :|:      |  ||  : :    ||:||||| | :  |:|  :|: :|:|
Db       2216 WLTLAGKGGMAAVSLPEARLRERIERFGQ----RLSVAAVNSPGTAAVAGDVDALRELLA 2271

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              :|||  :| : || ||   ||  :  ::| :   | |::|| | |||:|:|||  |||
Db       2272 ELTAEGIRAKPIPGVDTAGHSAQVDGLKEHLFEVLAPVSPRSSDIPFYSTVTGAPLDTER 2331

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||| ||||||   | || | | |:  |   |:| :|||:| | | |   | |   ||
Db       2332 LDAGYWYRNMREPVEFEKAVRALIADGYDLFLECNPHPMLAMSLDETLTDSGG-HGT--- 2387

Qy        300 VIMGTLRRGQGTLDHF 315
              :| |||| :|:   |
Db       2388 -VMHTLRRQKGSAKDF 2402


RESULT 11
Q93HI8
ID   Q93HI8      PRELIMINARY;      PRT;  3970 AA.
AC   Q93HI8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69199.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3970 AA;  412657 MW;  A93E583FAFA68C47 CRC64;
```

```
  Query Match             46.8%;  Score 758;  DB 2;  Length 3970;
  Best Local Similarity   52.8%;  Pred. No. 3.5e-42;
  Matches  167;  Conservative   34;  Mismatches  105;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||  ||| :: |||  :  || |  ||||||:  |||  |  |  |||
Db       2372 VFVFPGQGAQWAGMAAELLVSAPVFRARIEECEEALRPYVDWSLLAVLRQEPGTPPWDRV 2431

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |:|||:|||||||| |||| |||||| |||||| | | | ||| |||| | ::|
Db       2432 DVVQPALWAVMVSLAALWRSYGVEPDAVLGHSQGEIAAACVCGALDLADGARVVALRARA 2491

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
               :|||| | :||||          || |    | |::||:|||| ||||:|  :|: :|:
Db       2492 IGSTLAGRGGMVSVALPSAAARELIAGW----PGRVSVASVNGPVSTVVAGDPDALDELL 2547

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              |    || | ||||  :|:  :  ||  : : |  :||| | ||| |::||  ||| :
Db       2548 ARCADDGVRARRIPVDYASHTSQVAEIEGDLAAELAGLTPRSSAIPFFSTLTGAPLDTAK 2607

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||| |||:|:  || || | |  | ||   | |:|||||   :|:   | | |
Db       2608 LDAGYWYQNLRHTVEFEAATRAALDQGFTVFAEISPHPVLVPSVQDTIDDTGRT-----A 2662

Qy        300 VIMGTLRRGQGTLDHF 315
              |':|:|||  |    |
Db       2663 VAVGSLRRDDGGARRF 2678
```

RESULT 12
Q93HE5
ID Q93HE5 PRELIMINARY; PRT; 4498 AA.
AC Q93HE5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Modular polyketide synthase.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMBL; AB070945; BAB69243.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000515; BPD_transp.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW Phosphopantetheine.
SQ SEQUENCE 4498 AA; 468412 MW; 9D64DCCB088B4E3D CRC64;

```
  Query Match             46.8%;  Score 756.5;  DB 2;  Length 4498;
  Best Local Similarity   52.8%;  Pred. No. 5.2e-42;
  Matches  167;  Conservative   34;  Mismatches  106;  Indels    9;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              | |||||| ||  | |||   : |||  :  |  | : :|:| :  ||        | ||
Db       1530 VLVFPGQGSQWGAMARELYAGAPVFRARLEECAKALSAFVEWDLLDVLLSDDGGVLLSRV 1589

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:|||||| : || | ||:||| ||||||:|:| ||| |||||||| |||
Db       1590 DVVQPALFAVMVSLAALWEAHGVRPDAVVGHSQGEIAAAYVAGVLSLEDAARVVTLRSQA 1649

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | :|||  : ||: ||:  |  |    : :||:||| |||||:  |: :| |
Db       1650 IRALAGRGGMVSVGLSVDEVRPRLGRWGGD----VEIAAINGPTSTVVSGSAGALDELQA 1705

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  |||  ||||  :||  :  : | |:  | |||||   : |:|:||| |:|| ||
Db       1706 VPEAENVRTWRIPVDYASHSCHVDRLRENVMDLLAPITPRAGDLRFYSTVTGRRIDTTEL 1765

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || ||:||:  || ||   ||||  |  ||:| |||||  |:||       |:| |
Db       1766 DADYWFRNLRETVEFEAVTRLLLADGFDTFIECSPHPVLVAGVQETI----DSTHTPARA 1821

Qy        301 IMGTLRRGQGTLDHFL 316
              | |||||  | || ||
Db       1822 I-GTLRRHHGGLDSFL 1836
```

RESULT 13
Q93NX9
ID Q93NX9 PRELIMINARY; PRT; 9510 AA.
AC Q93NX9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AmphI.
GN AMPHI.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood B.M., Finnan S.M., Oliynyk M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
RT nodosus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF357202; AAK73501.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS50075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_6.
KW Phosphopantetheine.
SQ SEQUENCE 9510 AA; 992817 MW; 1A8E003A15D478CF CRC64;

```
  Query Match             46.7%;  Score 756;  DB 2;  Length 9510;
  Best Local Similarity   51.7%;  Pred. No. 1.5e-41;
  Matches  164;  Conservative   44;  Mismatches   99;  Indels   10;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||| :||| | || | :. | || | : ||::  ||| :  || |:||
```

```
Db       3851 VFVFPGQGSQWAGMGAQLLDESPVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERV 3910

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120

Db       3911 DVVQPASFAVMVSLAAVWRAQGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQA 3970

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179

Db       3971 IGRSLAGRGGMMSVALPVAEVEARL----EAPEGRVSVAAENGPRSSVVAGEPEALDELH 4026

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239

Db       4027 AQLTAEEIRARRVAVDYASHSPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTV 4086

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299

Db       4087 MDAGYWYRSLRGRVLFADAVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAG-VAGVAS- 4144

Qy        300 VIMGTLRRGQGTLDHFL 316

Db       4145 ---GTLRRDNGGLDRFL 4158


RESULT 14
Q9EWA1
ID   Q9EWA1      PRELIMINARY;      PRT;  9507 AA.
AC   Q9EWA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PimS2 protein.
GN   PIMS2.
OS   Streptomyces natalensis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=68242;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20547809; PubMed=11094342;
RA   Aparicio J.F., Fouces R., Mendes M.V., Olivera N., Martin J.F.;
RT   "A complex multienzyme system encoded by five polyketide synthase
RT   genes is involved in the biosynthesis of the 26-membered polyene
RT   macrolide pimaricin in Streptomyces natalensis.";
RL   Chem. Biol. 7:895-905(2000).
DR   EMBL; AJ278573; CAC20921.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine.
SQ   SEQUENCE   9507 AA;  994228 MW;  57BD80C8AD37E89F CRC64;

  Query Match             46.6%;  Score 754.5;  DB 2;  Length 9507;
  Best Local Similarity   51.4%;  Pred. No. 1.8e-41;
  Matches  163;  Conservative   43;  Mismatches  100;  Indels   11;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60

Db       3859 VFVFPGQGSQWAGMGARLLEESPVFAARLTECATALSEFVDWSLLDVLRQADGAPTLDRV 3918

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120

Db       3919 DVVQPASFAVMVSLAALWTSHGITPDAVVGHSQGEIAAAAVAGALSLEDAARVVALRSQA 3978

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179

Db       3979 IARGLAGTGGMLSVPLPAADVEQRLAAY-ED----LSIAAVNGPRSTVVSGATAPLDALQ 4033

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239

Db       4034 AELVGEDIRARRIAVDYASHSAQVERVRDELRTVLEPVRPRPAQVPFFSTVTGDWLDTTV 4093

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299

Db       4094 MDAEYWFTNLRRTVHFQPAIGELLAQGHHFFIEVSSHPVLSMGIQATAEEAG-----AAA 4148

Qy        300 VIMGTLRRGQGTLDHFL 316

Db       4149 AVLGTLRRDTGATDRFL 4165


RESULT 15
Q9L4X3
ID   Q9L4X3      PRELIMINARY;      PRT;  9477 AA.
AC   Q9L4X3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysI.
GN   NYSI.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71766.1; -.
DR   HSSP; P81989; 1QE5.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   9477 AA;  988132 MW;  0D2BCA5D7B265483 CRC64;

  Query Match             46.5%;  Score 753;  DB 2;  Length 9477;
  Best Local Similarity   53.0%;  Pred. No. 2.3e-41;
  Matches  168;  Conservative   34;  Mismatches  105;  Indels   10;  Gaps    3;
```

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAPAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  |||  :|||  |  ||  |  :   |  ||  |  :  |||:  |||     ||  |:||
Db       3895 VFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAALAEFTDWSLVDVLRGVVGAPSLERV 3954

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||||:||||||||:||  |  ||:|||  ||||||  |||  |||  |  ||||  |  |||
Db       3955 DVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQGEIAAAVVSGALSLRDGARVVALRSQA 4014

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
                  |||  |  ::|||  :  |  |  ||:  :       |::|||||||||  ||:|  ||:  |
Db       4015 IGRALAGRGGMMSVALSVDVLEPRLVEF----EGRVSVAAVNGPRSVVVAGEPEALDALH 4070

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              |  |||  :|  |  |  ||  :||  :  :  |  ::  |  :  ||  |  :||  |:|||  |||
Db       4071 ARLTADDIRARRIAVDYASHSHQVEDLHEELLEVLAELAPRTSEVPFFSTVTGDWLDTAR 4130

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              :||  ||:||:   |||  |  ||    : |||:|  ||||:|  :||   : |
Db       4131 MDAGYWFRNLRGRVRFADAVADLLAAEYRAFVEVSSHPVLSMAVQEAIDEAG-----VPA 4185

Qy        300 VIMGTLRRGQGTLDHFL 316
              |  |||||  ||  |  ||
Db       4186 VAAGTLRRDQGGTDRFL 4202
```

Search completed: June 17, 2003, 13:07:51
Job time : 15.6048 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 3.44608 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_29_344
Perfect score: 1618
Sequence: 1 VFVFPGQGPQWPGMGRELLD.....ADTVIMGTLRRGQGTLDHFL 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	776.5	48.0	3567	1	ERY2_SACER	Q03132 saccharopol
2	722	44.6	3491	1	ERY1_SACER	Q03131 saccharopol
3	707	43.7	3519	1	OL56_STRAT	Q07017 streptomyce
4	689	42.6	3172	1	ERY3_SACER	Q03133 saccharopol
5	463.5	28.6	2110	1	MCAS_MYCBO	Q02251 mycobacteri
6	410	25.3	1538	1	PPSB_MYCTU	Q10978 mycobacteri
7	398	24.6	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	366.5	22.7	1774	1	MSAS_PENPA	P22367 penicillium
9	265	16.4	317	1	FABD_BACSU	P71019 bacillus su
10	258	15.9	1986	1	WA_EMENI	Q03149 emericella
11	255	15.8	308	1	FABD_SALTY	O85140 salmonella
12	252	15.6	308	1	FABD_ECOLI	P25715 escherichia
13	247	15.3	2109	1	PKS1_ASPPA	Q12053 aspergillus
14	243.5	15.0	2181	1	STCA_EMENI	Q12397 emericella
15	217	13.4	293	1	FABD_SYNY3	P73242 synechocyst
16	213.5	13.2	312	1	FABD_HAEIN	P43712 haemophilus
17	205.5	12.7	2505	1	FAS_RAT	P12785 rattus norv
18	187	11.6	2511	1	FAS_CHICK	P12276 gallus gall
19	167.5	10.4	2504	1	FAS_HUMAN	P49327 homo sapien
20	125.5	7.8	2037	1	FAS1_CANAL	P34731 c fatty aci
21	123.5	7.6	463	1	YI02_MYCTU	O53951 mycobacteri
22	123	7.6	2073	1	FAS1_SCHPO	Q9uug0 s fatty aci
23	109	6.7	1914	1	STCK_EMENI	Q00706 emericella
24	108.5	6.7	920	1	PARC_SYNY3	P73077 synechocyst
25	107	6.6	2035	1	HFC1_HUMAN	P51610 homo sapien
26	104	6.4	426	1	G6PI_HALN1	Q9hnq6 halobacteri
27	103	6.4	547	1	MUTL_DEIRA	Q9rtr0 deinococcus
28	102.5	6.3	343	1	GPDA_RHILO	P58142 rhizobium l
29	101	6.2	494	1	MURC_MYCTU	O06225 mycobacteri
30	101	6.2	2090	1	HFC1_MESAU	P51611 mesocricetu
31	100.5	6.2	509	1	YD30_MYCTU	Q10641 mycobacteri
32	99	6.1	375	1	EPC2_RALSO	P52641 ralstonia s
33	97	6.0	852	1	SYQ_DEIRA	P56926 deinococcus
34	97	6.0	2051	1	FAS1_YEAST	P07149 s fatty aci
35	96.5	6.0	302	1	FABD_MYCTU	Q10501 mycobacteri
36	96.5	6.0	304	1	LDH_DEIRA	P50933 deinococcus
37	96	5.9	1074	1	CARB_SALTI	Q8z917 salmonella
38	96	5.9	1074	1	CARB_SALTY	P14846 salmonella
39	96	5.9	1108	1	CYGE_RAT	P51840 rattus norv
40	96	5.9	1286	1	IRBP_BOVIN	P12661 bos taurus
41	95.5	5.9	225	1	GRPE_STRCO	Q05562 streptomyce
42	95.5	5.9	408	1	YS92_MYCTU	Q10813 mycobacteri
43	95.5	5.9	1103	1	CYGD_HUMAN	Q02846 homo sapien
44	94.5	5.8	518	1	VL2_HPV12	P36748 human papil
45	94	5.8	375	1	EPC1_RALSO	P58600 ralstonia s

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             48.0%;  Score 776.5;  DB 1;  Length 3567;
  Best Local Similarity   50.0%;  Pred. No. 2.3e-48;
  Matches  158;  Conservative   44;  Mismatches  107;  Indels    7;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||:||||| || ||| ||| :| ||   :|||: : ||  || |  ||| :| |||||||
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:||| |||| |||| ||:||| ||||||||:| |:| |||::|   |:
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                :|:| | : :||    . :  |: || :    ||:|||||||| ||||   |:
Db        680 MRSLSGEGGMAAVALGEAAVRERLRPWQD----RLSVAAVNGPRSVVVSGEPGALRAFSE 735

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  | :| | | ||  :||| :  : | ::     | |||:|: |||:|    :|  ||
Db        736 DCAAEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTEL 795

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:  |||| |   | : |   |:|:|||||:   ::|    : :  |  | |
Db        796 DARYWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEA---VEEADGAEDAV 852

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|:| |  | |  ||
Db        853 VVGSLHRDGGDLSAFL 868
```

```
RESULT 2
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
```

```
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             44.6%;  Score 722;  DB 1;  Length 3491;
  Best Local Similarity   50.3%;  Pred. No. 2.1e-44;
  Matches  159;  Conservative   42;  Mismatches  103;  Indels   12;  Gaps    5;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||| || ||  ||| | ||  :: ||  || |  ||:: ||| |||:      ||
Db         57 VFVFPGQGAQWAGMAGELLGESRVFAAAMDACARAFEPVTDWTLAQVL-DSPEQS--RRV 113

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              :|||| |||| |||||||| || | ||:|||:||:||||| |    |||||   |||:
Db        114 EVVQPALFAVQTSLAALWRSFGVTPDAVVGHSIGELAAAHVCGAAGAADAARAAALWSRE 173

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                | | | : :|| : ||: |||| | :|    : :| |||||| :::|: | ||  |
Db        174 MIPLVGNGDMAAVALSADEIEPRIARWDDD----VVLAGVNGPRSVLLTGSPEPVARRVQ 229

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :|:|  || ::| | : |||  :  | | : | |    |  | :||::|:||| :|||||
Db        230 ELSAEGVRAQVINVSMAAHSAQVDDIAEGMRSALAWFAPGGSEVPFYASLTGGAVDTREL 289

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | || |:    |||: | |  |: || |||| ||||||   ||:      |  |::  |
Db        290 VADYWRRSFRLPVRFDEAIRSALEVGPGTFVEASPHPVLAAALQQTL----DAEGSSAAV 345

Qy        301 IMGTLRRGQGTLDHFL 316
              :  ||:|||| :  ||
Db        346 V-PTLQRGQGGMRRFL 360


RESULT 3
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1     ?       MODULE 5.
FT   DOMAIN        ?  3519       MODULE 6.
FT   DOMAIN       32   501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569   890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200  1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487  1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686  2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220  2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856  3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141  3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?     ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?     ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270  3519       THIOESTERASE.
FT   ACT_SITE    210   210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660   660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203  1249       NADP.
FT   BINDING    1524  1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859  1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311  2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859  2905       NADP.
FT   BINDING    3178  3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             43.7%;  Score 707;  DB 1;  Length 3519;
  Best Local Similarity   49.4%;  Pred. No. 2.6e-43;
  Matches  157;  Conservative   41;  Mismatches  106;  Indels   14;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||    ||||  || |:|  | |    |   ||: :||:   ||  | ||
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQ-GRDATVLGRV 2279

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |:|||:|||   ||  |||| ||:||| |||||| |:||||||| |||| | |:|
Db       2280 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 2339

Qy        121 QTTLAGTGALVSVAATPDELLP--RIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL 178
                :|| | :|||:       ||  |:    :    |::||||||| ||||||  :|: :|
Db       2340 IARIAGGGGMVSVS------LPAGRVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQALDEL 2393

Qy        179 VADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTR 238
```

```
              :|      || | :|||  :||  |  : : ::  |  |||: | :|| |:||   |||
Db       2394 LAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTT 2453

Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
              ||| ||: |:  ||||: |   |: ||   ||| ||||||  |:::   |       :
Db       2454 ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQ-----N 2508

Qy        299 TVIMGTLRRGQGTLDHFL 316
                ::|:||| :| ||  |
Db       2509 AAVLGSLRRDEGGLDRLL 2526


RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGPTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
```

```
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             42.6%;  Score 689;  DB 1;  Length 3172;
  Best Local Similarity   46.8%;  Pred. No. 4.6e-42;
  Matches  148;  Conservative   52;  Mismatches  102;  Indels   14;  Gaps    5;

Qy          3 VFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDV 62
              |||||| || || |:|| | || :|:| || | ||:||||:  :|      |  ||||||
Db        557 VFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDV 613

Qy         63 VQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQT 122
              ||| |||||:|||||||| |||| ||:||| |||||||:| |:| |||::| : |:
Db        614 VQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLR 673

Qy        123 TLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS-TVVSGAREAVADLVAD 181
              | | | : |     ::   ||  :       |::|:|||||  ||:|   : :|:|:
Db        674 RLGGQGGMASFGLGTEQAAERIGRPA----GALSIASVNGPRSVVVVAGESGPLDELIAE 729

Qy        182 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELD 241
              |   : | ||||  :||| : :: | ::: |  |:|  : :  :|: ||  :||  :|
Db        730 CEAEAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMD 789

Qy        242 AAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD--T 299
              |||| |:   |||: | | | : |   |||:|||||||:|::       |:   ||
Db        790 TAYWYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATL----DSALPADAGA 845

Qy        300 VIMGTLRRGQGTLDHF 315
              ::||||| :| |  |
Db        846 CVVGTLRRDRGGLADP 861


RESULT 5
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             28.6%;  Score 463.5;  DB 1;  Length 2110;
  Best Local Similarity   36.4%;  Pred. No. 7.1e-26;
  Matches  107;  Conservative   57;  Mismatches  117;  Indels   13;  Gaps    5;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |:|| ||| ||  || :|| :  ||  ::   |   |    :|| : :       |:|:
Db        534 VWVFSGQGSQWAAMGTQLLASEPVFAATIAKLEPVIAAESGFSVTEAITAQQTVTGIDK- 592

Qy         61 DVVQPTLFAVMISLAA-LWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQ 119
                ||| :||| ::||| : :: || | ||:|||:|| ||| |:| ||| |||||:   |:
Db        593 --VQPAVFAVQVALAATMEQTYGVRPGAVVGHSMGESAAAVVAGALSLEDAARVICRRSK 650

Qy        120 AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
                | :|| ||: ||     ::   :     |:   : |: |  |:|||: |  : | ||:
Db        651 LMTRIAGAGAMGSVELPAKQVNSELMARGIDD---VVVSVVASPQSTVIGGTSDTVRDLI 707

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              |      | | : ||| :||| :  | : : : |  | |   ::|::|:      | ||
Db        708 ARWEQRDVMAREVAVDVASHSPQVDPILDDLAAALADIAPMTPKVPYYSATL---FDPRE 764

Qy        240 ---LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDL 290
                 | |||  |: :||:|  | :  :: | : | ::||||:||  :::    |
Db        765 QPVCDGAYWVDNLRNTVQFAAAVQAAMEDGYRVFADVSPHPLLTHAVEQTGRSL 818


RESULT 6
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
```

```
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match           25.3%;  Score 410;  DB 1;  Length 1538;
  Best Local Similarity   34.1%;  Pred. No. 3.8e-22;
  Matches  109;  Conservative   56;  Mismatches  125;  Indels   30;  Gaps   10;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||: |:| || ||||:||     |  :|   |  |     :|:  |:    :  |::: 
Db        561 VFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLRDVIATGKELVGIEQ- 619

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
                :|  |  : ::|  |||| ||:|  |:|||:||:||| |:| |: |:  ||    ::
Db        620 --IQLGLIGMQLTLTELWRSYGVQPDLVIGHSMGEVAAAVVAGALTPAEGLRVTATRARL 677

Qy        121 QTTLAGTG--ALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL 178
               |:| |  ||: : |   | |         | | :: |   | || ||::|  | : :|
Db        678 MAPLSGQGGMALLGLDAAATEAL------IADYP-QVTVGIYNSPRQTVIAGPTEQIDEL 730

Qy        179 VADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTR 238
               :| : |       : ::|  |:| | |::  : | |  :|||   |    | |   | |:
Db        731 IARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGI-ISTTYADLHTQ 789

Qy        239 EL DAAYWYRNMSSTVRFEPAARLLLQQGP------KTFVEMSPHPVLTMGLQELAPDLG 291
              : || :|  || : |||: |   :   |         ||:|:| ||:||   | :|  |
Db        790 PIFDAEHWATNMRNPVRFQQA---IASAGSGADGAYHTFIEISAHPLLT---QAIADTLE 843

Qy        292 D----TTGTADTVIMGTLRR 307
              |    |   |  : :|||:|
Db        844 DAHRPTKSAAKYLSIGTLQR 863


RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             24.6%;  Score 398;  DB 1;  Length 1876;
  Best Local Similarity   33.6%;  Pred. No. 3.6e-21;
  Matches  107;  Conservative   53;  Mismatches  128;  Indels   30;  Gaps    9;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||: |:| || ||||:||     |  :|    |  |      :|:  || :  :  |::: 
Db        632 VFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLHDVLANGEELVGIEQ- 690

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
                :|  |  : ::|  || | || |  |:|||:||:||| |:| |: |:  ||    |: 
Db        691 --IQLGLIGMQLALTELWCSYGVRPDLVIGHSMGEVAAAVVAGALTPAEGLRVTATRSRL 748

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTE----DNPARLAVAAVNGPRSTVVSGAREAVA 176
              |:| | :          ||   || ||    | | :: :   | || ||::|  | : 
Db        749 MAPLSGQGGMA--------LLELDAPTTEALIADFP-QVTLGIYNSPRQTVIAGPTEQID 799

Qy        177 DLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD 236
              :|:| : |       : ::|  |:| | |::  : | |  :|||  |    | |  |
Db        800 ELIARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGI-ISTTYADLH 858

Qy        237 TREL-DAAYWYRNMSSTVRFEPAARLLLQQGP------KTFVEMSPHPVLTMGLQELAPD 289
              |: : || :|  || : |||: |   :   |         ||:|:| ||:||   | :
Db        859 TQPVFDAEHWATNMRNPVRFQQA---IASAGSGADGAYHTFIEISAHPLLT---QAIIDT 912

Qy        290 LGDTTGTADTVIMGTLRR 307
              |      |    :|||:|
Db        913 LHSAQPGARYTSLGTLQR 930


RESULT 8
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             22.7%;  Score 366.5;  DB 1;  Length 1774;
  Best Local Similarity   31.0%;  Pred. No. 6.6e-19;
  Matches  100;  Conservative   64;  Mismatches  132;  Indels   27;  Gaps    8;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |:|| | | ||| ||::|:  : ||  :::  :      :  |  ::|| : |    |||
Db        567 VWVFSGHGAQWPDMGKQLIH-NPVFFAAIQPLDELIQAEIGLSPIELLR-TGDFESSDRV 624

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
                ::    : : | |:|| :| |: | ||:|||:|||||: |:| || |: | :||  :
Db        625 QIL---TYVMQIGLSALLQSNGITPQAVIGHSVGEIAASVVAGALSPAEGALIVTRRALL 681

Qy        121 QTTLAGTGALVSV---AATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVAD 177
                : | | :: |   :|  :|:|          : : | ||  : | | ||:| :| ||:
Db        682 YRQVMGKGGMILVNLPSAETEEIL--------GSRSDLVVAIDSSPSSCVVAGDKELVAE 733

Qy        178 LVADLTAAQVRTRMIPVDVPAHSP----LMYAIEERVVSGLLPITPRPSRIPFHSSVTGG 233
                | |  |:|  :  |:  |||    |:  : : :   | |::|    :  :|:
Db        734 TAEALKARGVKTFTVKSDIAFHSPTLNGLVDPLRDVLAETLSPVSP---NVKLYSTALAD 790

Qy        234 RLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDT 293
                  |  ||  || : ||   | :  :: | : |:|:|  |||::  : |   |
Db        791 PRGQDLRDVEYWAGNMVNRVRLTSAVKAAVEDGYRLFLEVSTHPVVSHSINETLMD---- 846
```

```
Qy        294 TGTADTVIMGTLRRGQGTLDHFL 316
                |  |  :: || | : |  | |
Db        847 AGMEDFAVIPTLLRKKPTEKHIL 869


RESULT 9
FABD_BACSU
ID   FABD_BACSU     STANDARD;      PRT;   317 AA.
AC   P71019; O34463;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96326321; PubMed=8759840;
RA   Morbidoni H.R., de Mendoza D., Cronan J.E. Jr.;
RT   "Bacillus subtilis acyl carrier protein is encoded in a cluster of
RT   lipid biosynthesis genes.";
RL   J. Bacteriol. 178:4794-4800(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98195738; PubMed=9534248;
RA   Foulger D., Errington J.;
RT   "A 28 kbp segment from the spoVM region of the Bacillus subtilis 168
RT   genome.";
RL   Microbiology 144:801-805(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U59433; AAC44306.1; -.
DR   EMBL; Y13937; CAA74249.1; -.
DR   EMBL; Z99112; CAB13463.1; -.
DR   HSSP; P25715; 1MLA.
DR   SubtiList; BG11836; fabD.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; Complete proteome.
FT   ACT_SITE     91     91       BY SIMILARITY.
FT   ACT_SITE    201    201       BY SIMILARITY.
FT   CONFLICT    124    131       VPAGEGAM -> GCRLAKEQW (IN REF. 1).
SQ   SEQUENCE   317 AA;  34035 MW;  191AE828B1C91D7F CRC64;

  Query Match             16.4%;  Score 265;  DB 1;  Length 317;
  Best Local Similarity   28.3%;  Pred. No. 2.1e-12;
  Matches   83;  Conservative   50;  Mismatches  130;  Indels   30;  Gaps    8;

Qy          2 FVFPGQGPQWPGMGRELLD----ASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGL 57
              |:||||| |: |||:|| :    |  :| |:    |   :  :           ||  |
Db          6 FLFPGQGSQFIGMGKELYEQVPAAKRLFDEADETLETKLSSLI---------FEGDAEEL 56

Qy         58 DRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLW 117
                 || |    |::   ::  |: |    |||||| :|   :| ||  ||   |
Db         57 TLTYNAQPALLTTSIAVLEKFKESGITPDFTAGHSLGEYSALVAAGALSFKDAVYTVRKR 116

Qy        118 SQ--AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAV 175
              :   :   || ||: ::     | | ::     :    : :| :| | |   |:||  : |
Db        117 GEFMNEAVPAGEGAMAAILGMDAEALKQVTDKVTEEGNLVQLANLNCPGQIVISGTAKGV 176

Qy        176 ADLVADLTAAQVRTRMIPVDV--PAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGG 233
               :| ::|       | ||::|  | || ||    |::   |      : : :|  |:|:
Db        177 -ELASELAKENGAKRAIPLEVSGPFHSELMKPAAEKLKEVLDACDIKDADVPVISNVSAD 235

Qy        234 ----RLDTRE--LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLT 280
                  : | :|  ::  |      | |||| :   |: :|  ||:|: |  ||:
Db        236 VMTEKADIKEKLIEQLY------SPVRFEESINKLIAEGVTTFIEIGPGKVLS 282


RESULT 10
WA_EMENI
ID   WA_EMENI       STANDARD;      PRT;  1986 AA.
AC   Q03149;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Conidial green pigment synthase (EC 2.3.1.-).
GN   WA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93101122; PubMed=1465094;
RA   Mayorga M.E., Timberlake W.E.;
RT   "The developmentally regulated Aspergillus nidulans wA gene encodes a
RT   polypeptide homologous to polyketide and fatty acid synthases.";
RL   Mol. Gen. Genet. 235:205-212(1992).
CC   -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC       INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC       BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC       ASEXUAL SPORES (CONIDIA).
```

```
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Conidial green pigment biosynthesis.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65866; CAA46695.1; -.
DR   PIR; S28353; S28353.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN      529    582       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   DOMAIN      991   1024       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   DOMAIN     1650   1719       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1772   1841       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    548    548       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE   1001   1001       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1682   1682       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1804   1804       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1986 AA;  216634 MW;  74EF0940FF40EE9A CRC64;

  Query Match             15.9%;  Score 258;  DB 1;  Length 1986;
  Best Local Similarity   27.2%;  Pred. No. 5.8e-11;
  Matches   90;  Conservative   54;  Mismatches  139;  Indels   48;  Gaps   10;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEA--------AFAPYVDWSVEQVLRDSPD 53
              |:| ||| |:  ||::| :    || ::   :         :  | || |:  :   ||
Db        912 FLFTGQGAQYAAMGKQLYEDCSHFRSAIEHLDCISQGQDLPSILPLVDGSL-PLSELSP- 969

Qy         54 APGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLAD---- 109
                  |||     | ::|:: | | |: |  |||||||: || : :| || :|
Db        970 -------VVVQLGTTCVQMALSSFWASLGITPSFVLGHSLGDFAAMNAAGVLSTSDTIYA 1022

Qy        110 -AARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVV 168
               |   | : |    || |:::: |   |:   :      |      :| :| |  ||:
Db       1023 CGRRAQLLTERCQ---PGTHAMLAIKAPLVEVKQLLNEKVHD------MACINSPSETVI 1073

Qy        169 SGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHS 228
              || : :: :|    :   ::: :: |   ||  :  | | :   |  ||    :|| |
Db       1074 SGPKSSIDELSRACSEKGLKSTILTVPYAFHSAQVEPILEDLEKALQGITFNKPSVPFVS 1133

Qy        229 SVTG---GRLDTRELDAAYWYRNMSSTVR----FEPAARLLLQQGPKTFVEMSPHPVLT- 280
              :: |       :  |:| |  |:   ||      ||       ::|: || | :
Db       1134 ALLGEVITEAGSNILNAEYLVRHCRETVNFLSAFEAVRNAKLGGDQTLWLEVGPHTVCSG 1193

Qy        281 MGLQELAPDLGDTTGTADTVIMGTLRRGQGT 311
              |    | |          | | :||| : |
Db       1194 MVKATLGP---------QTTTMASLRRDEDT 1215


RESULT 11
FABD_SALTY
ID   FABD_SALTY     STANDARD;      PRT;   308 AA.
AC   O85140;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD OR STM1194.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=98317265; PubMed=9642179;
RA   Zhang Y., Cronan J.E. Jr.;
RT   "Transcriptional analysis of essential genes of the Escherichia coli
RT   fatty acid biosynthesis gene cluster by functional replacement with
RT   the analogous Salmonella typhimurium gene cluster.";
RL   J. Bacteriol. 180:3295-3303(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF044668; AAC38649.1; -.
DR   EMBL; AE008752; AAL20123.1; -.
DR   HSSP; P25715; 1MLA.
DR   StyGene; SG10637; fabD.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE     91     91       BY SIMILARITY.
FT   ACT_SITE    200    200       BY SIMILARITY.
SQ   SEQUENCE   308 AA;  32274 MW;  B3565B45341A611A CRC64;

  Query Match             15.8%;  Score 255;  DB 1;  Length 308;
  Best Local Similarity   28.6%;  Pred. No. 1.1e-11;
  Matches   90;  Conservative   50;  Mismatches  141;  Indels   34;  Gaps   11;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| |  ||  |:     :  |:     ||    |  |:: |       |  |::
Db          5 FVFPGQGSQSVGMLAEMAANYPIVEETFAEASAALG-YDLWALTQ----QGPAEELNKTW 59

Qy         62 VVQPTLFAVMISLAALWRSQGVE-PCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQ- 119
                 || |    ::|  :|: || : |  : |||||| :|   :| :: ||| |:| :  :
Db         60 QTQPALLTASVALWRVWQQQGGKMPALMAGHSLGEYSALVCAGVINFADAVRLVEMRGKF 119

Qy        120 -AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAV--NGPRSTVVSGAREAVA 176
               :    ||| : ::    |      ||   |::     |: |  | |   |::| :|||
Db        120 MQEAVPEGTGGMSAIIGLDD---ASIAKACEESAEGQVVSPVNFNSPGQVVIAGHKEAVE 176
```

```
Qy        177 DLVADLTAAQVRTRM-IPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRL 235
              |   ||  :  : :|| ||:|  ||    :::   |  ||     :|  ::|
Db        177 RAGAACKAAGAKRALPLPVSVPSHCALMKPAADKLAVELAKITFSAPTVPVVNNV----- 231

Qy        236 DTR-ELDAA----YWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDL 290
              | : | |||       | : : |::  :   :  || :   |: |  |||
Db        232 DVKCETDAAAIRDALVRQLYNPVQWTKSVEFIAAQGVEHLYEVGPGKVLT---------- 281

Qy        291 GDTTGTADTVIMGTL 305
              | |    ||:   |
Db        282 GLTKRIVDTLTASAL 296


RESULT 12
FABD_ECOLI
ID   FABD_ECOLI     STANDARD;      PRT;   308 AA.
AC   P25715;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD OR TFPA OR B1092.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12;
RX   MEDLINE=92183950; PubMed=1339356;
RA   Magnuson K., Oh W., Larson T.J., Cronan J.E. Jr.;
RT   "Cloning and nucleotide sequence of the fabD gene encoding malonyl
RT   coenzyme A-acyl carrier protein transacylase of Escherichia coli.";
RL   FEBS Lett. 299:262-266(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92234941; PubMed=1314802;
RA   Verwoert I.I., Verbree E.C., van der Linden K.H., Nijkamp H.J.,
RA   Stuitje A.R.;
RT   "Cloning, nucleotide sequence, and expression of the Escherichia coli
RT   fabD gene, encoding malonyl coenzyme A-acyl carrier protein
RT   transacylase.";
RL   J. Bacteriol. 174:2851-2857(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95214624; PubMed=7700236;
RA   Bouquin N., Tempete M., Holland I.B., Seror S.J.;
RT   "Resistance to trifluoroperazine, a calmodulin inhibitor, maps to the
RT   fabD locus in Escherichia coli.";
RL   Mol. Gen. Genet. 246:628-637(1995).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [6]
RP   SEQUENCE OF 288-308 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92210530; PubMed=1556094;
RA   Cronan J.E. Jr., Rawlings M.;
RT   "The gene encoding Escherichia coli acyl carrier protein lies within
RT   a cluster of fatty acid biosynthetic genes.";
RL   J. Biol. Chem. 267:5751-5754(1992).
RN   [7]
RP   SEQUENCE OF 1-11.
RC   STRAIN=K12 / W3110;
RA   Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA   Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA   Hochstrasser D.F.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
RN   [8]
RP   SEQUENCE OF 1-10.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=95286570; PubMed=7768883;
RA   Serre L., Verbree E.C., Dauter Z., Stuitje A.R., Derewenda Z.S.;
RT   "The Escherichia coli malonyl-CoA:acyl carrier protein transacylase
RT   at 1.5-A resolution. Crystal structure of a fatty acid synthase
RT   component.";
RL   J. Biol. Chem. 270:12961-12964(1995).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M87040; AAA23742.1; -.
DR   EMBL; Z11565; CAA77658.1; -.
DR   EMBL; M84991; AAA23738.1; -.
DR   EMBL; AE000210; AAC74176.1; -.
DR   EMBL; D90745; BAA35900.1; -.
DR   PIR; S20443; S20443.
DR   PIR; B41856; B41856.
DR   PDB; 1MLA; 25-JAN-95.
DR   SWISS-2DPAGE; P25715; COLI.
DR   EcoGene; EG11317; fabD.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; 3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   ACT_SITE     91     91
FT   ACT_SITE    200    200
SQ   SEQUENCE   308 AA;  32286 MW;  F5901043D92FED8E CRC64;

  Query Match            15.6%;  Score 252;  DB 1;  Length 308;
  Best Local Similarity  27.9%;  Pred. No. 1.8e-11;
  Matches   88;  Conservative   49;  Mismatches  144;  Indels   34;  Gaps   10;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              |||||| |  ||  ::  :  :  |:     ||   |  |:: |       | |::
Db          5 FVFPGQGSQTVGMLADMAASYPIVEETFAEASAALG-YDLWALTQ----QGPAEELNKTW 59
```

```
Qy          62 VVQPTLFAVMISLAALWRSQGVE-PCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQ- 119
               || |    ::|  :|: || : |  : |||||| :|   :| :  ||| |:| :  :
Db          60 QTQPALLTASVALYRVWQQQGGKAPAMMAGHSLGEYSALVCAGVIDFADAVRLVEMRGKF 119

Qy         120 -AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAV--NGPRSTVVSGAREAVA 176
                :    ||||: ::    |     ||   |:       |: | | |   |::| :|||
Db         120 MQEAVPEGTGAMAAIIGLDD---ASIAKACEEAAEGQVVSPVNFNSPGQVVIAGHKEAVE 176

Qy         177 DLVADLTAAQVRTRM-IPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSV----- 230
                |   ||  :  : :|| ||:|  ||    :::   |  ||     :|  ::|
Db         177 RAGAACKAAGAKRALPLPVSVPSHCALMKPAADKLAVELAKITFNAPTVPVVNNVDVKCE 236

Qy         231 TGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDL 290
               | |       |      | : : |::  :   :  || :   |: |  |||
Db         237 TNGDAIRDAL-----VRQLYNPVQWTKSVEYMAAQGVEHLYEVGPGKVLT---------- 281

Qy         291 GDTTGTADTVIMGTL 305
               | |    ||:   |
Db         282 GLTKRIVDTLTASAL 296


RESULT 13
PKS1_ASPPA
ID   PKS1_ASPPA     STANDARD;      PRT;  2109 AA.
AC   Q12053;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aflatoxin biosynthesis polyketide synthase (PKS).
GN   PKSL1.
OS   Aspergillus parasiticus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2999;
RX   MEDLINE=96042102; PubMed=7592391;
RA   Feng G.H., Leonard T.J.;
RT   "Characterization of the polyketide synthase gene (pksL1) required
RT   for aflatoxin biosynthesis in Aspergillus parasiticus.";
RL   J. Bacteriol. 177:6246-6254(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: Aflatoxin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L42766; AAC41675.1; -.
DR   EMBL; L42765; AAC41674.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Transferase; Acyltransferase; Phosphopantetheine;
KW   Multifunctional enzyme.
FT   DOMAIN      374    805       BETA-KETOACYL SYNTHASE.
FT   DOMAIN        ?      ?       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1714   1785       ACYL CARRIER (ACP).
FT   DOMAIN        ?   2109       THIOESTERASE.
FT   ACT_SITE    543    543       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    993    993       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1746   1746       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1937   1937       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2109 AA;  230715 MW;  CB701372A16D8551 CRC64;

  Query Match             15.3%;  Score 247;  DB 1;  Length 2109;
  Best Local Similarity   28.5%;  Pred. No. 3.9e-10;
  Matches   94;  Conservative   56;  Mismatches  118;  Indels   62;  Gaps   16;

Qy           1 VFVFPGQGPQWPGMGRELLDASDVFRESVR-----ACEAAFAPYV------DWSVEQVLR 49
               :| | ||| |:  ||::: ||   ||| :      |    |  ::        ||::
Db         902 LFAFTGQGSQYATMGKQVYDAYPSFREDLEKFDRLAQSHGFPSFLHVCTSPKGDVEEM-- 959

Qy          50 DSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLAD 109
                 ||      |||  :  : ::|  |  | |: |   :|||||| || : :| || :|
Db         960 ----AP-----VVVQLAITCLQMALTNLMTSFGIRPDVTVGHSLGEFAALYAAGVLSASD 1010

Qy         110 AARVVTLWSQAQTTL-----AGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPR 164
                 || |  |     |       || |:::| |||: |    : | :|:     || :|||
Db        1011 ---VVYLVGQRAELLQERCQRGTHAMLAVKATPEAL----SQWIQDHDCE--VACINGPE 1061

Qy         165 STVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAI---EERVVSG-------L 214
                ||:||  : ||::   :|   ::  :: :    ||  :  |     | :  |       |
Db        1062 DTVLSGTTKNVAEVQRAMTDNGIKCTLLKLPFAFHSAQVQPILDDFEALAQGATFAKPQL 1121

Qy         215 LPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQG---PKTFV 271
               | ::|   |   |         :   :  :|  ::   ||    | |   ::|    || |
Db        1122 LILSPL-LRTEIH--------EQGVVTPSYVAQHCRHTVDMAQALRSAREKGLIDDKTLV 1172

Qy         272 -EMSPHPVLTMGLQELAPDLGDTTGTADTV 300
                |: | |::: |: ::   |||   |  |:
Db        1173 IELGPKPLIS-GMVKMT--LGDKISTLPTL 1199


RESULT 14
STCA_EMENI
ID   STCA_EMENI     STANDARD;      PRT;  2181 AA.
AC   Q12397;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN   STCA OR PKSST.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
RX   MEDLINE=96202293; PubMed=8643646;
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 4;
RX   MEDLINE=95370159; PubMed=7642507;
RA   Yu J.-H., Leonard T.J.;
RT   "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT   novel type I polyketide synthase.";
RL   J. Bacteriol. 177:4792-4800(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
```

```
CC   -!- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34740; AAC49191.1; -.
DR   EMBL; L39121; AAA81586.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW   Multifunctional enzyme.
FT   DOMAIN      383    814       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      884   1209       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1706   1777       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1830   1901       ACYL CARRIER (ACP) 2.
FT   DOMAIN        ?   2181       THIOESTERASE.
FT   ACT_SITE    552    552       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    978    978       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1738   1738       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1862   1862       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2028   2028       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2181 AA;  238831 MW;  5A3E5712AA9AD942 CRC64;

  Query Match             15.0%;  Score 243.5;  DB 1;  Length 2181;
  Best Local Similarity   27.8%;  Pred. No. 7.3e-10;
  Matches   90;  Conservative   56;  Mismatches  153;  Indels   25;  Gaps    8;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |    ||| |: ||| :| ::   ||  :   :      |   | :|    |   :: :
Db        885 VIACSGQGSQYTGMGWQLYNSYPTFRSDLERFDQLARSYGFPSFLEVYTSKPVGDSMEDL 944

Qy         61 --DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWS 118
                :||  | :: ::|  |  | |::| ||:|||||| || ::|| || ||   :| : :
Db        945 LPVIVQLALVSLEMALGNLLGSFGLKPSAVIGHSLGEYAALYISGVLSAADTLYLVGMRA 1004

Qy        119 Q--AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVA 176
              :   :    || |:::| |:|  |   :|   | |     ||  |||  ||:||  : |
Db       1005 KLLQERCQRGTHAMLAVRASPVTLCEVLA---ESN---CEVACHNGPNDTVLSGPLKEVM 1058

Qy        177 DLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD 236
              :|   |:| .::  :: :    ||  :  | |   :    :|    :||  | :    :|
Db       1059 NLQNSLSATGIKGTLLKLPFAFHSAQVQPILEEFKNVARGVTFHKPQIPVLSPLLVKVID 1118

Qy        237 TR-ELDAAYWYRNMSSTVRFEPA---AR---------LLLQQGPKTFVEMSPHPVLTMGL 283
               :  :|  |  |:    |:         ||          :::  |||  :      |
Db       1119 EKGTVDPVYLARHCREPVKMVSVLEHARDQHIITDRTIVIDVGPKALMAGMIKTTLDKDT 1178

Qy        284 QELAPDLGDTTGT--ADTVIMGTL 305
                | || :     : | |:|||
Db       1179 SSALPTLGPSLDVWKSLTNILGTL 1202


RESULT 15
FABD_SYNY3
ID   FABD_SYNY3     STANDARD;      PRT;   293 AA.
AC   P73242;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD OR SLR2023.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90904; BAA17269.1; -.
DR   HSSP;\P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; Complete proteome.
FT   ACT_SITE     88     88       BY SIMILARITY.
FT   ACT_SITE    188    188       BY SIMILARITY.
SQ   SEQUENCE   293 AA;  31491 MW;  31E4ABA59ECDB377 CRC64;

  Query Match             13.4%;  Score 217;  DB 1;  Length 293;
  Best Local Similarity   27.8%;  Pred. No. 5.9e-09;
  Matches   83;  Conservative   46;  Mismatches  136;  Indels   34;  Gaps   10;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              :|||||| |  ||| :||  : : :|  :  |      : |||  | :   |   |
Db          5 WVFPGQGSQAVGMGVDLLSTA-IAKEKYQQAE----EILGWSV--VEKCQGDEASLALTQ 57

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQ 121
                || |: :   || | | :| :|  | ||||||  :| : :|     |   ::|   |:
Db         58 NTQPCLYVIEAILADLLRDKGFQPDYVAGHSLGEYSALYAAGVFDFATGLQLVKQRSEVM 117

Qy        122 TTLAG--TGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
               : :|    ||:    |  :|  :   ||     : :|  | |    |:||    |  ::
Db        118 ASASGGMMAALMKFDQT--QLQQALTDNTE-----VVLANDNSPEQVVISGTVAGVEAIL 170

Qy        180 ADLTAAQVRTRMIPVDVPA--HSPLMYAIEERVVSGLLPITPRPSRIPFHSSV------T 231
              |:: |      | :|: |    ||  |     :      |       : :|  |:|
Db        171 ANVKA----RRAVPLKVSGAFHSSFMAQPSQSFAQTLTACHFNDATVPVLSNVDPSPTQN 226

Qy        232 GGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDL 290
              | ||  : :       : |: :||:       | : |   : |: |  |||    :   |||
Db        227 GDRLKEKLI------QQMTGSVRWRETMVNLGEIGATDYWEVGPGKVLTGLCKRTCPDL 279


Search completed: June 17, 2003, 13:02:31
Job time : 5.44608 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 7.05626 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_29_344
Perfect score: 1618
Sequence: 1 VFVFPGQGPQWPGMGRELLD.....ADTVIMGTLRRGQGTLDHFL 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	817.5	50.5	2124	2	T28658	polyketide synthas
2	773	47.8	4151	2	G70944	probable polyketid
3	771.5	47.7	3573	2	S23070	erythronolide synt
4	722	44.6	3491	2	T43231	probable 6-deoxyer
5	707	43.7	3519	2	S43048	polyketide synthas
6	699	43.2	1937	2	T03224	probable polyketid
7	697.5	43.1	5069	2	T17464	rifamycin polyketi
8	696.5	43.0	4613	2	T17409	polyketide synthas
9	689.5	42.6	2126	2	H70621	probable polyketid
10	689	42.6	3172	2	S22012	erythronolide synt
11	689	42.6	3178	2	S13595	6-deoxyerythronoli
12	681.5	42.1	3739	2	T17410	polyketide synthas
13	674	41.7	1562	2	T17411	polyketide synthas
14	673.5	41.6	1763	2	T17465	rifamycin polyketi
15	667.5	41.3	1346	2	T17412	polyketide synthas
16	661	40.9	1728	2	T17466	rifamycin polyketi
17	658	40.7	4735	2	T17463	rifamycin polyketi
18	649.5	40.1	3413	2	T17467	rifamycin polyketi
19	638	39.4	7576	2	T17428	FK506 polyketide s
20	600	37.1	2723	2	T03221	probable polyketid
21	595	36.8	2100	2	T03223	probable polyketid
22	578	35.7	8563	2	T30226	polyketide synthas
23	566	35.0	6260	2	T30228	polyketide synthas
24	553	34.2	10223	2	T30225	polyketide synthas
25	539	33.3	6420	2	T30283	polyketide synthas
26	516.5	31.9	1570	2	AC2012	hypothetical prote
27	502.5	31.1	1582	2	E70876	probable polyketid
28	502	31.0	2108	2	H70819	probable polyketid
29	479	29.6	2126	2	E70522	probable polyketid
30	467.5	28.9	2111	2	A70668	mycocerosate synth
31	466.5	28.8	2118	2	S72705	mycocerosate synth
32	463.5	28.6	2110	2	B44110	mycocerosate synth
33	460.5	28.5	2116	2	C86926	probable mycoceros
34	457.5	28.3	1293	2	T30871	orsellinic acid sy
35	441.5	27.3	1827	2	B70984	probable polyketid
36	440	27.2	2458	2	T17420	probable polyketid
37	437	27.0	1587	2	AB2012	hypothetical prote
38	436	26.9	2297	2	T34918	polyketide synthas
39	432.5	26.7	1795	2	D97312	hypothetical prote
40	429	26.5	395	2	S18954	fix23-2 protein -
41	427	26.4	1017	2	B70985	probable polyketid
42	417	25.8	635	2	S73017	polyketide synthas
43	417	25.8	1446	2	S73013	polyketide synthas
44	417	25.8	1540	2	H87203	polyketide synthas
45	417	25.8	1871	2	A87204	polyketide synthas

ALIGNMENTS

RESULT 1
T28658
polyketide synthase - Sorangium cellulosum (fragment)
C;Species: Sorangium cellulosum
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C;Accession: T28658
R;Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J.J.; Ligon, J.M.
J. Bacteriol. 177, 3673-3679, 1995
A;Title: A Sorangium cellulosum (myxobacterium) gene cluster for the biosynthesis of th
inomycetes.
A;Reference number: Z20499; MUID:95325306; PMID:7601830
A;Accession: T28658
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2124 <SCH>
A;Cross-references: EMBL:U24241; NID:g868150; PID:g868151; PIDN:AAA79984.1
A;Experimental source: strain So ce26
C;Function:
A;Description: involved in soraphen A biosynthesis
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;335-406/Domain: acyl carrier protein homology <ACP>
F;370/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 50.5%; Score 817.5; DB 2; Length 2124;
Best Local Similarity 53.5%; Pred. No. 3.5e-53;
Matches 169; Conservative 42; Mismatches 96; Indels 9; Gaps 2;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||| || ||   |||:| |||  : ||| | ||:|:||:  |||    || ||||
Db        969 VFVFPGQGSQWEGMALSLLDSSPVFRAQLEACERALAPHVEWSLLAVLRRDEGAPSLDRV 1028

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:|||||||| |||| ||:||| |||||| |:| ||| ||||:  | |:|
Db       1029 DVVQPALFAVMVSLAALWRSLGVEPAAVVGHSQGEIAAAFVAGALSLEDAARIAALRSKA 1088

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||:|| ||: :|     :|   :||| :    ||::|||| ||:|:|||   |:  |:
Db       1089 LTTVAGNGAMAAVELGASDLQTYLAPWGD----RLSIAAVNSPRATLVSGEPAAIDALID 1144

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              ||||||  | : ||  :||  | |::: : :||  | ||   :| :|:||| |||  ||
Db       1145 SLTAAQVFARKVRVDYASHSAQMDAVQDELAAGLANIAPRTCELPLYSTVTGTRLDGSEL 1204

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | ||||||:  || |  |  ||  | : |||:|||||||: |:|          | |
Db       1205 DGAYWYRNLRQTVLFSSATERLLDDGHRFFVEVSPHPVLTLALRETC-----ERSPLDPV 1259

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|::|| :| |   |
Db       1260 VVGSIRRDEGHLARLL 1275
```

```
RESULT 2
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

  Query Match             47.8%;  Score 773;  DB 2;  Length 4151;
  Best Local Similarity   51.6%;  Pred. No. 1.8e-49;
  Matches  163;  Conservative   45;  Mismatches  102;  Indels    6;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||| || ||| |||| :  | : : ||  ||| :|||:  ||| :| |||||||
Db        559 VFVFPGQGSQWLGMGIELLDTAPAFAQQIDACAEAFAEFVDWSLVDVLRGAPGAPGLDRV 618

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:||| ||:|  | | ||:||| ||||||:|:| ||| ||||||||  |:
Db        619 DVVQPVLFAVMVSLAELWKSVAVHPDAVIGHSQGEIAAAYVAGALSLRDAARVVTLRSKL 678

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 ||| | :||:|   |:    :||: :   |:::| |||| : |||   |: :|:|
Db        679 LAGLAGPGGMVSIACGADQARDLLAPFGD----RVSIAVVNGPSAVVVSGEVGALEELIA 734

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :  ::||| | ||  :||  : ||    :   |  | || :|  | |:||| |||| |
Db        735 VCSTKELRTRRIEVDYASHSVEVEAIRGPLAEALSGIEPRSTRTVFFSTVTGNRLDTAGL 794

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:  || |: | |   :|| :||:| |||| |  |::|       | | :: :
Db        795 DADYWYRNVRQTVLFDQAVRNACEQGYRTFIESSPHPALITGVEETF--AACTDGDSEAI 852

Qy        301 IMGTLRRGQGTLDHFL 316
              :: || || | |  ||
Db        853 VVPTLGRGDGGLHRFL 868


RESULT 3
S23070
erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase II
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C;Accession: S23070; S22011; S23205
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S23070
A;Molecule type: DNA
A;Residues: 1-3573 <BEV1>
A;Cross-references: EMBL:X62569
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22011
A;Molecule type: DNA
A;Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,1
3479,'DH',3480-3572 <BEV2>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44448.1; PID:g581651
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of t
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23205
A;Molecule type: protein
A;Residues: 2-12,'XXX' <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
A;Start codon: GTG
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ort-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; mult
F;52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;1404-1475/Domain: acyl carrier protein homology <ACP1>
F;1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;3420-3493/Domain: acyl carrier protein homology <ACP2>
F;1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             47.7%;  Score 771.5;  DB 2;  Length 3573;
  Best Local Similarity   49.7%;  Pred. No. 2e-49;
  Matches  157;  Conservative   44;  Mismatches  108;  Indels    7;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||:||||| || ||| ||| :| ||   :|||: : ||  || |  ||| :| |||||||
Db        561 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 620

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:||| |||| |||| ||:||| ||||| ||:| |:| |||::|    |:
Db        621 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAEHVAGALTLEDAAKLVVGRSRL 680

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                :|:| | : :||     :  |: || :    ||:||||||||| ||||   |:
Db        681 MRSLSGEGGMAAVALGEAAVRERLRPWQD----RLSVAAVNGPRSVVVSGEPGALRAFSE 736

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  |  :| | | ||  :||| :  : | ::     | |||:|: |||:|    :|  ||
Db        737 DCAAEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTEL 796

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:  ||||  |   | : |   |:|:|||||:   ::|    : :  |  | |
Db        797 DARYWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEA---VEEADGAEDAV 853

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|:| |  | |  ||
Db        854 VVGSLHRDGGDLSAFL 869


RESULT 4
T43231
```

probable 6-deoxyerythronolide B synthase, module 1/2 - Saccharopolyspora erythraea
C;Species: Saccharopolyspora erythraea
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000
C;Accession: T43231
R;Donadio, S.; Staver, M.J.; McAlpine, J.B.; Swanson, S.J.; Katz, L.
Science 252, 675-679, 1991
A;Title: Modular organization of genes required for complex polyketide biosynthesis.
A;Reference number: Z22354; MUID:91220065; PMID:2024119
A;Accession: T43231
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3491 <DON>
A;Cross-references: EMBL:M63676; NID:g152691; PID:g152692; PIDN:AAA26493.1
C;Genetics:
A;Gene: eryA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;57-336/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;416-483/Domain: acyl carrier protein homology <ACP1>
F;526-926/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1031-1317/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;1886-1957/Domain: acyl carrier protein homology <ACP2>
F;2001-2400/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2508-2791/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;3332-3403/Domain: acyl carrier protein homology <ACP3>
F;1921/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 44.6%; Score 722; DB 2; Length 3491;
Best Local Similarity 50.3%; Pred. No. 1e-45;
Matches 159; Conservative 42; Mismatches 103; Indels 12; Gaps 5;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  ||   |||   | ||   ::  ||   ||  |   ||::  |||  |||:      ||
Db         57 VFVFPGQGAQWAGMAGELLGESRVFAAAMDACARAFEPVTDWTLAQVL-DSPEQS--RRV 113

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              :||||  ||||   ||||||||  ||  |  ||:|||:||:|||||  |     |||||    |||:
Db        114 EVVQPALFAVQTSLAALWRSFGVTPDAVVGHSIGELAAAHVCGAAGAADAARAAALWSRE 173

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 | | | : :|| : ||: |||| | :|    : :| ||||||  :::|: | ||  |
Db        174 MIPLVGNGDMAAVALSADEIEPRIARWDDD----VVLAGVNGPRSVLLTGSPEPVARRVQ 229

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :|:|  || ::| | : |||  :  | | : | |    |  | :||::|:||| :|||||
Db        230 ELSAEGVRAQVINVSMAAHSAQVDDIAEGMRSALAWFAPGGSEVPFYASLTGGAVDTREL 289

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | || |:    |||: | |  |: || |||| ||||||   ||:      |  |::  |
Db        290 VADYWRRSFRLPVRFDEAIRSALEVGPGTFVEASPHPVLAAALQQTL----DAEGSSAAV 345

Qy        301 IMGTLRRGQGTLDHFL 316
              :  ||:|||| :  ||
Db        346 V-PTLQRGQGGMRRFL 360
```

RESULT 5
S43048
polyketide synthase type I - Streptomyces antibioticus
N;Contains: acyl carrier protein; acyltransferase; ketoacylsynthase; ketoreductase; thio
C;Species: Streptomyces antibioticus
C;Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 26-May-2000
C;Accession: S43048; S41729
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, February 1993
A;Reference number: S43048
A;Accession: S43048
A;Molecule type: DNA
A;Residues: 1-3519 <SWA>
A;Cross-references: EMBL:L09654; NID:g153407; PIDN:AAA19695.1; PID:g153408
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A;Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyketi
A;Reference number: S41729; MUID:94150470; PMID:8107683
A;Accession: S41729
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1683-3238,3273-3303,'K',3305-3407,'T',3409-3462,'Y',3464-3516,'E',3518-3519
A;Cross-references: EMBL:L09654
C;Genetics:
A;Start codon: GTG
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;570-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1202-1381/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1489-1560/Domain: acyl carrier protein homology <ACP1>
F;1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3143-3214/Domain: acyl carrier protein homology <ACP2>
F;3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 43.7%; Score 707; DB 2; Length 3519;
Best Local Similarity 49.4%; Pred. No. 1.4e-44;
Matches 157; Conservative 41; Mismatches 106; Indels 14; Gaps 4;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  ||   ||||  ||  |:|  | |   |   ||: :||:   ||  | ||
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQ-GRDATVLGRV 2279

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  |:|||:|||  ||  |||| ||:||| ||||||  |:|||||| |||| | |:|
Db       2280 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 2339

Qy        121 QTTLAGTGALVSVAATPDELLP--RIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL 178
                 :|| | :|||:      ||  |:    :    |::|||||| ||||||  :|: :|
Db       2340 IARIAGGGGMVSVS------LPAGRVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQALDEL 2393

Qy        179 VADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTR 238
              :|      || | :|||  :||  |  : : ::  |  |||: | :|| |:||   |||
Db       2394 LAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTT 2453

Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
              ||| ||: |:  ||||: |   |: ||   ||| ||||||  |:::    |      :
Db       2454 ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQ-----N 2508

Qy        299 TVIMGTLRRGQGTLDHFL 316
              ::|:||| :| ||  |
Db       2509 AAVLGSLRRDEGGLDRLL 2526
```

RESULT 6
T03224
probable polyketide synthase module 4 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03224
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03224
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1937 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38064.1; PID:g2624951
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;549-827/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

Query Match 43.2%; Score 699; DB 2; Length 1937;

```
  Best Local Similarity   48.7%;  Pred. No. 2.6e-44;
  Matches  153;  Conservative   39;  Mismatches  100;  Indels   22;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  ||||  | ::| |||:||   :  | | :||||   ||   |  | ||||
Db        549 VFVFPGQGSQWIGMGRGLWESSAVFRDSVLRTDRALAEFVDWSAAAVLAGEPGTPDLDRV 608

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||  ||::||  :|||  ||||  ||:||| ||||||||:|||||  ||||||  |  |:|
Db        609 DVVQPVLFTVMVALANVWRSYGVEPAAVVGHSQGEIAAAHVAGGLSLRDAARVVALRSRA 668

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              | |||  | :|  |  :   :   :  |       ||:|| ||||  : ||||   |: :|:
Db        669 LTALAGEGGMVVVQRSVAYVENLLRQW----EGRLSVAVVNGPEAVVVSGQVAALEELL- 723

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  : |  |  : ||  :||  :  |||::   |  : |  ||:|  |:|    :||  :
Db        724 ---ATEDRARRVAVDYASHSAQVERIEEKLTRTLTDVQPMTSRVPLFSTVERDWIDTASM 780

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              |  |||||:  || |: |   | : |   |||:|||||||  ::|  |          : |
Db        781 DTGYWYRNLRQTVWFDDAVGRLSEAGFDVFVEVSPHPVLTASVRESCP---------EAV 831

Qy        301 IMGTLRRGQGTLDH 314
              : |||||     ||
Db        832 VTGTLRR-----DH 840


RESULT 7
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             43.1%;  Score 697.5;  DB 2;  Length 5069;
  Best Local Similarity   48.7%;  Pred. No. 1.1e-43;
  Matches  155;  Conservative   41;  Mismatches  107;  Indels   15;  Gaps    5;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |:|||||| || ||||||||:|:||   :  || |  :||||:  |||   |   ||||
Db        570 VWVFPGQGTQWAGMGRELLDSSEVFAARIAECETALGRWVDWSLTDVLRGEADL--LDRV 627

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||||: |||:| | |||| ||:||| |||||| ||| ||| |||:|| | |||
Db        628 DVVQPASFAVMVGLAAVWASLGVEPEAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQA 687

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
               :||| | : ||| : ::   |: ||      |: |||||| | |::|  ||: : :
Db        688 IAASLAGRGGMASVALSEEDATARLEPWA----GRVEVAAVNGPTSVVIAGDAEALDEAL 743

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRL-DTR 238
              .  |   || | : ||  :|:  : |  : :   |  |  :   :||:|:|||| : |
Db        744 DALDDQGVRIRRVAVDYASHTRHVEAARDALAEMLGGIRAQAPEVPFYSTVTGGWVEDAG 803

Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
               ||  |||||:   ||| ||   |::|| : |||:| ||||   : ||  |        :
Db        804 VLDGGYWYRNLRRQVRFGPAVAELIEQGHRVFVEVSAHPVLVQPINELVDD-------TE 856
```

```
Qy        299 TVIMGTLRRGQGTLDHFL 316
              |: |||||  | |   |
Db        857 AVVTGTLRREDGGLRRLL 874


RESULT 8
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

  Query Match             43.0%;  Score 696.5;  DB 2;  Length 4613;
  Best Local Similarity   49.2%;  Pred. No. 1.2e-43;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              |||||||  || |||  ||||:|  ||   ::   |||| :||||||:| |:|  :|  || |:|||
Db        634 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 693

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:   || | ||:||| |||||||:|:| ||| ||||||||  |::
Db        694 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 753

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              |||  | ::|:| : | :|  |:|  :        |:|||||||  :|||||     : :|
Db        754 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 808

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:||||  :||  :  ||  :   |  ::|:  |:|| |:: |  :  |
Db        809 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 868

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| ||||||  |        |  || |
Db        869 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 920

Qy        300 VIMGTLRRGQGTLDHFL 316
                |||| |  |  :
Db        921 ----TLRRDNGGQDRLV 933


RESULT 9
H70621
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70621
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
```

A;Residues: 1-2126 <COL>
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06632.1; PID:g1839013
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks7
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;563-845/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1694-1875/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1976-2051/Domain: acyl carrier protein homology <ACP2>

Query Match 42.6%; Score 689.5; DB 2; Length 2126;
Best Local Similarity 47.1%; Pred. No. 1.5e-43;
Matches 148; Conservative 46; Mismatches 115; Indels 5; Gaps 2;

```
Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| :|   : || | :  || |   :|:||:  ||| :| ||||||||
Db        564 FVFPGQGAQWLGMGAQLCATAPVFAEHIHRCERALREHVEWSLLDVLRGAPGAPGLDRVD 623

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQ 121
              |||| |:|||:||| |||| || | ||:||| |||||||:|:| ||| ||| || | |:
Db        624 VVQPALWAVMVSLAELWRSVGVVPDAVIGHSQGEIAAAYVAGALSLRDAAAVVALRSRLL 683

Qy        122 TTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVAD 181
                | | | :||:|    :     : | :    || :|||||  | |::|  :|| :|:
Db        684 VRLGGAGGMVSLACGQPQAEKLASQWGD----RLNIAAVNGVSSVVLAGETDAVTELMQR 739

Qy        182 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELD 241
              |  :| | | ||  :||  : || | ::: |  | || | : | |:|||  :||  ::
Db        740 CEAEGIRARRIDVDYASHSAQVDAIREELIAALRGIEPRTSTVAFFSTVTGELMDTAGVN 799

Qy        242 AAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVI 301
              | ||||::   |:|| | |     | : ||| ||||||  |::|   | |   | : ::
Db        800 AEYWYRSIRQPVQFERAVRNAFDGGYRVFVESSPHPVLIAGIEETLVDC-DRGATGEPIV 858

Qy        302 MGTLRRGQGTLDHF 315
              : || |  | :  |
Db        859 IPTLGRDDGGVGRF 872
```

RESULT 10
S22012
erythronolide synthase (EC 2.3.1.94) III - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase III; protein A
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S22012; S34784; S23206
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22012
A;Molecule type: DNA
A;Residues: 1-3172 <BEV1>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44449.1; PID:g46979
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S34784
A;Molecule type: DNA
A;Residues: 1-715,717-728,'C',729-886,888-894,'QR',895-910,'E',912-1276,'RR',1279-1284,'
EV2>
A;Cross-references: EMBL:X62569
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23206
A;Molecule type: protein
A;Residues: 2-11,'XXX',15 <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; mult
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;555-835/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1117-1297/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1392-1463/Domain: acyl carrier protein homology <ACP1>
F;1511-1914/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2022-2300/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2557-2734/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2819-2890/Domain: acyl carrier protein homology <ACP2>
F;2959-3153/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 42.6%; Score 689; DB 2; Length 3172;
Best Local Similarity 46.8%; Pred. No. 2.7e-43;
Matches 148; Conservative 52; Mismatches 102; Indels 14; Gaps 5;

```
Qy          3 VFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDV 62
              |||||| || || |:||  | || :|:| || | ||:||||:  :|     |  ||||||
Db        557 VFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDV 613

Qy         63 VQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQT 122
              ||| ||||:|||||||| |||| ||:||| ||||||||:| |:| |||::| : |:
Db        614 VQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLR 673

Qy        123 TLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS-TVVSGAREAVADLVAD 181
              | | | : |      ::   ||  :       |::|:||||||  ||:|    : :|:|:
Db        674 RLGGQGGMASFGLGTEQAAERIGRFA----GALSIASVNGPRSVVVVAGESGPLDELIAE 729

Qy        182 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELD 241
              |   : | ||||  :||| : :: | ::: |  |:|  : :  :|: ||  :||  :|
Db        730 CEAEAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMD 789

Qy        242 AAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD--T 299
              |||| |:   |||: | | | : |   |||:|||||||:|::       |:   ||
Db        790 TAYWYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATL----DSALPADAGA 845

Qy        300 VIMGTLRRGQGTLDHF 315
              ::||||| :| |  |
Db        846 CVVGTLRRDRGGLADF 861
```

RESULT 11
S13595
6-deoxyerythronolide B synthase - Saccharopolyspora erythraea
N;Alternate names: protein A
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S13595
R;Cortes, J.; Haydock, S.F.; Roberts, G.A.; Bevitt, D.J.; Leadlay, P.F.
Nature 348, 176-178, 1990
A;Title: An unusually large multifunctional polypeptide in the erythromycin-producing p
A;Reference number: S13595; MUID:91043075; PMID:2234082
A;Accession: S13595
A;Molecule type: DNA
A;Residues: 1-3178 <COR>
A;Cross-references: EMBL:X56107; NID:g46975; PIDN:CAA39583.1; PID:g46976
A;Experimental source: strain NRRL 2338
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
-carrier-protein] S-malonyltransferase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; multifunctional
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

```
F;553-833/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1123-1303/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1398-1469/Domain: acyl carrier protein homology <ACP1>
F;1517-1920/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2028-2306/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2563-2740/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2825-2896/Domain: acyl carrier protein homology <ACP2>
F;2965-3159/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

  Query Match            42.6%;  Score 689;  DB 2;  Length 3178;
  Best Local Similarity   46.8%;  Pred. No. 2.8e-43;
  Matches  148;  Conservative   52;  Mismatches  102;  Indels   14;  Gaps    5;

Qy          3 VFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDV 62
              |||||| || || |:||  | || :|:| || | ||:||||:  :|      |  ||||||
Db        555 VFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDV 611

Qy         63 VQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQT 122
              ||| |||||:|||||||| |||| ||:||| |||||||:| |:| |||::| : |:
Db        612 VQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLR 671

Qy        123 TLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS-TVVSGAREAVADLVAD 181
              | | | : |     ::   ||  :       |::|:|||||  ||:|    : :|:|:
Db        672 RLGGQGGMASFGLGTEQAAERIGRFA----GALSIASVNGPRSVVVVAGESGPLDELIAE 727

Qy        182 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELD 241
              |   : | ||||  :||| : :: | ::: |  |:|  : :  :|: ||  :|| . :|
Db        728 CEAEAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMD 787

Qy        242 AAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD--T 299
              |||| |:   |||: | | | : |   |||:|||||||:|::      |:   ||
Db        788 TAYWYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATL----DSALPADAGA 843

Qy        300 VIMGTLRRGQGTLDHF 315
              ::|||||  :| |  |
Db        844 CVVGTLRRDRGGLADF 859


RESULT 12
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

  Query Match            42.1%;  Score 681.5;  DB 2;  Length 3739;
  Best Local Similarity   48.9%;  Pred. No. 1.2e-42;
  Matches  152;  Conservative   42;  Mismatches   98;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||| |  |  ::  |||| |||||||:| |:| :| || |:|||
Db       2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| ||||||||  |::
Db       2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:|     :| |:: :       |:|||||| :|||||    : :|
Db       2189 GAHLAGQGGMLSLALSEAAVVERLAGFD-----GLSVAAVNGPTATVVSGDPTQIQELAQ 2243

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  ::|| |:: |  :    |
Db       2244 ACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPAL 2303

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              | |||||:   | | ||   |   :|   |||:| |||||| | |         ||
Db       2304 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPE------TVTG---- 2353

Qy        300 VIMGTLRRGQG 310
               :|||||  |
Db       2354 --LGTLRRDNG 2362


RESULT 13
T17411
polyketide synthase III - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17411
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17411
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1562 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C;Genetics:
A;Gene: pikAIII
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1403-1474/Domain: acyl carrier protein homology <ACP>

  Query Match            41.7%;  Score 674;  DB 2;  Length 1562;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-42;
  Matches  149;  Conservative   41;  Mismatches  102;  Indels   18;  Gaps    4;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||| :|  |  |:  || | : |||||:| |:|  | || |||||
Db        565 FVFPGQGTQWAGMGAELLDSSPEFAASMAECETALSRYVDWSLEAVVRQEPGAPTLDRVD 624

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  |: | ||:||| ||||||:|:| |:| ||||||||  |::
Db        625 VVQPVTFAVMVSLAKVWQHHGITPQAVVGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 684

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:|     :| |:: :       |:|||||| :|||||    : :|
Db        685 AAHLAGKGGMISLALDEAAVLKRLSDFD-----GLSVAAVNGPTATVVSGDPTQIEELAR 739

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||: :   |  : |:   :|| |:: |  :    |
Db        740 TCEADGVRARIIPVDYASHSRQVEIIEKELAEVLAGLAPQAPHVPFFSTLEGTWITEPVL 799

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | |||||:   | | ||   |   |:|:| |||||| | |         ||
Db        800 DGTYWYRNLRHRVGFAPAVETLAVDGFTHFIEVSAHPVLTMTLPE------TVTG----- 848

Qy        301 IMGTLRRGQG 310
               :|||||  ||
Db        849 -LGTLRREQG 857


RESULT 14
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
```

```
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>

  Query Match             41.6%;  Score 673.5;  DB 2;  Length 1763;
  Best Local Similarity   47.8%;  Pred. No. 1.9e-42;
  Matches  152;  Conservative   39;  Mismatches  108;  Indels   19;  Gaps    5;

Qy          1 VFVPPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |:|||||| || ||||||||||| || | :: | ||   : |||:  |||   |   || |
Db        555 VWVFPGQGTQWVGMGRELLDASPVFAERIKECAAALDQWTDWSLLDVLRGDGD---LDSV 611

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              :|:||  ||||: |||:| | || | ||:||| |||||| ||| |:| |||:|| | |||
Db        612 EVLQPACFAVMVGLAAVWESAGVRPDAVVGHSQGEIAAACVSGALTLDDAAKVVALRSQA 671

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
               |:| | : ||| : ||   |:  |      |: ||||||| | |::|  :|: : :
Db        672 IAARLSGRGGMASVALSEDEANARLGLW----DGRIEVAAVNGPASVVIAGDAQALDEAL 727

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRL-DTR 238
              |    || | : ||  :|:  :  | : :   |  || :   :|| |:|||| : |
Db        728 EVLAGDGVRVRQVAVDYASHTRHVEDIRDTLAETLAGITAQAPDVPFRSTVTGGWVRDAD 787

Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
              ||  |||||: : ||| ||   ||:||   |||:| ||||   : ||           |
Db        788 VLDGGYWYRNLRNQVRFGPAVAELLEQGHGVFVEVSAHPVLVQPISEL----------TD 837

Qy        299 TVIMGTLRRGQGTLDHFL 316
               |: |||||  | |   |
Db        838 AVVTGTLRRDDGGLRRLL 855


RESULT 15
T17412
polyketide synthase IV - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17412
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17412
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1346 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800837; PIDN:AAC69332.1
C;Genetics:
A;Gene: pikAIV
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;945-1016/Domain: acyl carrier protein homology <ACP>

  Query Match             41.3%;  Score 667.5;  DB 2;  Length 1346;
  Best Local Similarity   47.6%;  Pred. No. 3.9e-42;
  Matches  148;  Conservative   45;  Mismatches   99;  Indels   19;  Gaps    6;

Qy          2 FVPPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:|  |  ::  || | :||||||:| |:| :| || |||||
Db        562 FVFPGQGTQWAGMGAELLDSSPEFAAAMAECETALSPYVDWSLEAVVRQAPSAPTLDRVD 621

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
                 ||||  ||||:||| :|:  |: | ||:||| ||||||:|:| |:| ||||||| |::
Db        622 VVQPVTFAVMVSLAKVWQHHGITPEAVIGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 681

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
               ||| | ::|:| : :    ||      :|   |::|||||| :|||||    : :|
Db        682 AAHLAGKGGMISLALSEEATRQRI-----ENLHGLSIAAVNGPTATVVSGDPTQIQELAQ 736

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
               |  :| |:|||| :||  :  ||  :   |  ::|:  ::|| |:: |  :    |
Db        737 ACEADGIRARIIPVDYASHSAHVETIENELADVLAGLSPQTPQVPFFSTLEGTWITEPAL 796

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
               |  |||||:   | | ||   |  :|   |:|:| |||||| |    ||    || |
Db        797 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTL----PD--KVTGLA-- 848

Qy        300 VIMGTLRRGQG 310
                ||||  |
Db        849 ----TLRREDG 855


Search completed: June 17, 2003, 13:12:19
Job time : 8.05626 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 9.84594 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_29_344
Perfect score: 1618
Sequence: 1 VFVPPGQGPQWPGMGRELLD.....ADTVIMGTLRRGQGTLDHFL 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	819	50.6	344	9	US-09-735-056-33	Sequence 33, Appl
2	714.5	44.2	3816	9	US-09-808-880-3	Sequence 3, Appli
3	712	44.0	4150	9	US-09-808-880-2	Sequence 2, Appli
4	707	43.7	3519	9	US-09-808-880-4	Sequence 4, Appli
5	696.5	43.0	4551	9	US-09-793-708-1	Sequence 1, Appli
6	696.5	43.0	4613	9	US-09-860-846-31	Sequence 31, Appl
7	696.5	43.0	4613	9	US-09-988-384B-31	Sequence 31, Appl
8	696.5	43.0	4613	9	US-09-836-821-31	Sequence 31, Appl
9	696.5	43.0	4613	10	US-09-861-289-31	Sequence 31, Appl
10	696.5	43.0	11877	9	US-09-860-846-6	Sequence 6, Appli
11	696.5	43.0	11877	9	US-09-836-821-6	Sequence 6, Appli
12	696.5	43.0	11877	10	US-09-861-289-6	Sequence 6, Appli
13	696.5	43.0	12199	9	US-09-988-384B-6	Sequence 6, Appli
14	681.5	42.1	3739	9	US-09-860-846-33	Sequence 33, Appl
15	681.5	42.1	3739	9	US-09-988-384B-33	Sequence 33, Appl
16	681.5	42.1	3739	9	US-09-836-821-33	Sequence 33, Appl
17	681.5	42.1	3739	9	US-09-793-708-2	Sequence 2, Appli
18	681.5	42.1	3739	10	US-09-861-289-33	Sequence 33, Appl
19	674	41.7	1562	9	US-09-860-846-35	Sequence 35, Appl
20	674	41.7	1562	9	US-09-988-384B-35	Sequence 35, Appl
21	674	41.7	1562	9	US-09-836-821-35	Sequence 35, Appl
22	674	41.7	1562	9	US-09-793-708-3	Sequence 3, Appli
23	674	41.7	1562	10	US-09-861-289-35	Sequence 35, Appl
24	667.5	41.3	1346	9	US-09-860-846-37	Sequence 37, Appl
25	667.5	41.3	1346	9	US-09-988-384B-37	Sequence 37, Appl
26	667.5	41.3	1346	9	US-09-836-821-37	Sequence 37, Appl
27	667.5	41.3	1346	9	US-09-793-708-4	Sequence 4, Appli
28	667.5	41.3	1346	10	US-09-861-289-37	Sequence 37, Appl
29	667	41.2	345	9	US-09-735-056-34	Sequence 34, Appl
30	627.5	38.8	3798	9	US-10-014-717-6	Sequence 6, Appli
31	583.5	36.1	7257	9	US-10-014-717-5	Sequence 5, Appli
32	574.5	35.5	1832	9	US-10-014-717-4	Sequence 4, Appli
33	441.5	27.3	1827	9	US-09-712-363-261	Sequence 261, App
34	411	25.4	2439	9	US-10-014-717-7	Sequence 7, Appli
35	404.5	25.0	343	9	US-09-735-056-32	Sequence 32, Appl
36	399.5	24.7	1616	9	US-09-712-363-262	Sequence 262, App
37	398	24.6	1421	9	US-10-014-717-2	Sequence 2, Appli
38	395	24.4	5215	9	US-09-860-846-2	Sequence 2, Appli
39	395	24.4	5215	9	US-09-988-384B-2	Sequence 2, Appli
40	395	24.4	5215	9	US-09-836-821-2	Sequence 2, Appli
41	395	24.4	5215	10	US-09-861-289-2	Sequence 2, Appli
42	386.5	23.9	1402	9	US-09-712-363-166	Sequence 166, App
43	383	23.7	328	9	US-09-735-056-31	Sequence 31, Appl
44	310	19.2	1610	9	US-09-738-626-6666	Sequence 6666, Ap
45	274.5	17.0	2563	9	US-09-836-705-46	Sequence 46, Appl

ALIGNMENTS

```
RESULT 1
US-09-735-056-33
; Sequence 33, Application US/09735056
; Publication No. US20030013662A1
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/735,056
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  344 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  No. US20030013662A1e
US-09-735-056-33

  Query Match             50.6%;  Score 819;  DB 9;  Length 344;
  Best Local Similarity   54.0%;  Pred. No. 4.2e-61;
  Matches  168;  Conservative   38;  Mismatches   95;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  || ||   ||: |  || :  :|:||  ||: |||  |||  |||| ||||
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  || :|:||||:||: |||| ||:||| ||||||||:| ||| |:||:| | |:|
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 ||| | :|:|    :|| ||:  | :    |||||||| | |  |:|  ||:|:|||
Db        125 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 180

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
                 ||    |  | || ||   ||| : |:   ::  | |: |||: |||:|:|||| ||  |
Db        181 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTE 240

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
                 ||| ||||||   | || | | |:  |   |:| ||||:| : |::   | |      |
Db        241 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 295

Qy        300 VIMGTLRRGQG 310
                ::|||||  |
Db        296 AVLGTLRRRHG 306


RESULT 2
US-09-808-880-3
; Sequence 3, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

  Query Match             44.2%;  Score 714.5;  DB 9;  Length 3816;
  Best Local Similarity   50.3%;  Pred. No. 6.1e-51;
  Matches  160;  Conservative   37;  Mismatches  106;  Indels   15;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  || ||   ||||  || |:|  | |    |   ||: :|||       | ||

Db        567 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLRGGEAV--LGRV 624

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |:|||:|||  ||  |||| ||:||| |||||| |:|||||| |||| | |:|
Db        625 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 684

Qy        121 QTTLAGTGALVSVAATPDELLP--RIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL 178
                 :|| | :|||:      ||  |:    :    ||:|||||| ||||||  :|: :|
Db        685 IARIAGGGGMVSVS------LPAGRVRTMLDTYGGRLSVAAVNGPSSTVVSGDAQALDEL 738

Qy        179 VADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTR 238
                 :|      || | :|||  :||  |  : : ::  |  |||: | :|| |:||   |||
Db        739 LAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTT 798

Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
                 ||| ||: |:  ||||: |   |: ||   ||| ||||||  |:::          | ||
Db        799 ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTL-----DTVEAD 853

Qy        299 TVIMGTLRRGQGTLDHFL 316
                | :|:||| :| |  ||
Db        854 AVALGSLRRDEGGLGRFL 871


RESULT 3
US-09-808-880-2
; Sequence 2, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-2

  Query Match             44.0%;  Score 712;  DB 9;  Length 4150;
  Best Local Similarity   50.3%;  Pred. No. 1.1e-50;
  Matches  160;  Conservative   37;  Mismatches  107;  Indels   14;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  || ||   ||||  || |:|  | |    |   ||: :|||   :|  | ||
Db       1581 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLRGGGEAV-LGRV 1639

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |:|||:|||  ||  |||| ||:||| |||||| |:|||||| |||| | |:|
Db       1640 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 1699

Qy        121 QTTLAGTGALVSVAATPDELLP--RIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL 178
                 :|| | :|||:      ||  |:    |:   |::|||||| ||||||  :|: :|
Db       1700 IARIAGGGGMVSVS------LPAGRVRTMLEEFDGRVSVAAVNGPSSTVVSGDVQALDEL 1753

Qy        179 VADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTR 238
                 :|      || | :|||  :||  |  : : ::  |  | |  : :|| |:||   |||
Db       1754 LAGCEREGVRARRVPVDYASHSAQMDQLRDDLLEALATIVPTSANVPFFSTVTADWLDTT 1813
```

```
Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
              ||| ||: |:  |||| | |: ||  ||| |||||| |: |       | ||
Db       1814 ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGITETL-----DTFDAD 1868

Qy        299 TVIMGTLRRGQGTLDHFL 316
              | : :||| :| || ||
Db       1869 AVALSSLRRDEGGLDRFL 1886


RESULT 4
US-09-808-880-4
; Sequence 4, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 3519
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-4

  Query Match           43.7%;  Score 707;  DB 9;  Length 3519;
  Best Local Similarity  49.4%;  Pred. No. 2.3e-50;
  Matches  157;  Conservative   41;  Mismatches  106;  Indels   14;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||   |||| || |:|  | |    |   ||: :||:   ||  | ||
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQ-GRDATVLGRV 2279

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |:|||:|||  ||  |||| ||:||| |||||| |:||||||| |||| | |:|
Db       2280 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 2339

Qy        121 QTTLAGTGALVSVAATPDELLP--RIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL 178
                :|| | :|||:      ||  |:    :    |::|||||| ||||||  :|: :|
Db       2340 IARIAGGGMVSVS------LPAGRVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQALDEL 2393

Qy        179 VADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTR 238
              :|      || | :||| :||  |  : : ::  |  |||: | :|| |:||   |||
Db       2394 LAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTT 2453

Qy        239 ELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTAD 298
              ||| ||: |:  |||| |   |: ||  ||| ||||||  |:::    |       :
Db       2454 ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQ-----N 2508

Qy        299 TVIMGTLRRGQGTLDHFL 316
              ::|:||| :| ||  |
Db       2509 AAVLGSLRRDEGGLDRLL 2526


RESULT 5
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-1

  Query Match           43.0%;  Score 696.5;  DB 9;  Length 4551;
  Best Local Similarity  49.2%;  Pred. No. 2.5e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||| :| ||  ::  |||| :||||||:| |:| :| || |:|||
Db        572 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 631

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              |||  |||| ||| :|:  || | ||:||| |||||| :| ||| |||||||| |::
Db        632 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 691

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:| : | :|. |:| :       |:|||||| :|||||    : :|
Db        692 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 746

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :    |
Db        747 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 806

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| |         | || |
Db        807 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 858

Qy        300 VIMGTLRRGQGTLDHFL 316
                 ||||  |  |  :
Db        859 ----TLRRDNGGQDRLV 871


RESULT 6
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
```

```
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-31

  Query Match             43.0%;  Score 696.5;  DB 9;  Length 4613;
  Best Local Similarity   49.2%;  Pred. No. 2.5e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:| ||  ::  |||| :||||||:| |:| :| || |:|||
Db        634 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 693

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |::
Db        694 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 753

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :|
Db        754 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 808

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :    |
Db        809 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 868

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| |       |  || |
Db        869 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 920

Qy        300 VIMGTLRRGQGTLDHFL 316
                  ||||  |  |  :
Db        921 ----TLRRDNGGQDRLV 933


RESULT 7
US-09-988-384B-31
; Sequence 31, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-31

  Query Match             43.0%;  Score 696.5;  DB 9;  Length 4613;
  Best Local Similarity   49.2%;  Pred. No. 2.5e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:| ||  ::  |||| :||||||:| |:| :| || |:|||
Db        634 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 693

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |::
Db        694 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 753

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :|
Db        754 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 808

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :    |
Db        809 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 868

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| |       |  || |
Db        869 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 920

Qy        300 VIMGTLRRGQGTLDHFL 316
                  ||||  |  |  :
Db        921 ----TLRRDNGGQDRLV 933


RESULT 8
US-09-836-821-31
; Sequence 31, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-31

  Query Match             43.0%;  Score 696.5;  DB 9;  Length 4613;
  Best Local Similarity   49.2%;  Pred. No. 2.5e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:| ||  ::  |||| :||||||:| |:| :| || |:|||
Db        634 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 693

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |::
Db        694 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 753

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :|
Db        754 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 808

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :    |
Db        809 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 868

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| |       |  || |
Db        869 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 920

Qy        300 VIMGTLRRGQGTLDHFL 316
                  ||||  |  |  :
Db        921 ----TLRRDNGGQDRLV 933


RESULT 9
US-09-861-289-31
; Sequence 31, Application US/09861289
; Patent No. US20020110897A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-31
```

```
  Query Match             43.0%;  Score 696.5;  DB 10;  Length 4613;
  Best Local Similarity   49.2%;  Pred. No. 2.5e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:| ||  ::  |||| :||||||:| |:| :| || |:||| 
Db        634 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 693

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |:: 
Db        694 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 753

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :| 
Db        754 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 808

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :   | 
Db        809 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 868

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   :|   |||:| |||||| |       |  || | 
Db        869 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 920

Qy        300 VIMGTLRRGQGTLDHFL 316
                |||| |  |  :
Db        921 ----TLRRDNGGQDRLV 933
```

```
RESULT 10
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6
```

```
  Query Match             43.0%;  Score 696.5;  DB 9;  Length 11877;
  Best Local Similarity   49.2%;  Pred. No. 8.8e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:| ||  ::  |||| :||||||:| |:| :| || |:||| 
Db        970 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 1029

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |:: 
Db       1030 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 1089

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :| 
Db       1090 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 1144

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :   | 
Db       1145 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 1204

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   :|   |||:| |||||| |       |  || | 
Db       1205 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 1256

Qy        300 VIMGTLRRGQGTLDHFL 316
                |||| |  |  :
Db       1257 ----TLRRDNGGQDRLV 1269
```

```
RESULT 11
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-6
```

```
  Query Match             43.0%;  Score 696.5;  DB 9;  Length 11877;
  Best Local Similarity   49.2%;  Pred. No. 8.8e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| ||||:| ||  ::  |||| :||||||:| |:| :| || |:||| 
Db        970 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 1029

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |:: 
Db       1030 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 1089

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :| 
Db       1090 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 1144

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :   | 
Db       1145 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 1204

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   :|   |||:| |||||| |       |  || | 
Db       1205 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 1256
```

```
Qy        300 VIMGTLRRGQGTLDHFL 316
                ||||  |  |  :
Db       1257 ----TLRRDNGGQDRLV 1269


RESULT 12
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             43.0%;  Score 696.5;  DB 10;  Length 11877;
  Best Local Similarity   49.2%;  Pred. No. 8.8e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||||:| ||  ::  |||| :||||||:| |:| :| || |:|||
Db        970 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 1029

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:|.||| |||||||| |::
Db       1030 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 1089

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :|
Db       1090 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 1144

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :    |
Db       1145 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 1204

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| |        |  || |
Db       1205 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 1256

Qy        300 VIMGTLRRGQGTLDHFL 316
                ||||  |  |  :
Db       1257 ----TLRRDNGGQDRLV 1269


RESULT 13
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             43.0%;  Score 696.5;  DB 9;  Length 12199;
  Best Local Similarity   49.2%;  Pred. No. 9.1e-49;
  Matches  156;  Conservative   42;  Mismatches  100;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||||:| ||  ::  |||| :||||||:| |:| :| || |:|||
Db       1292 FVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVD 1351

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |::
Db       1352 VVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 1411

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:| : | :| |:| :       |:||||||| :|||||    : :|
Db       1412 AAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTATVVSGDPVQIEELAR 1466

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||  :||  :  ||  :   |  ::|:  |:|| |:: |  :    |
Db       1467 ACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFSTLEGAWITEPVL 1526

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| |        |  || |
Db       1527 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP------GTVTGLA-- 1578

Qy        300 VIMGTLRRGQGTLDHFL 316
                ||||  |  |  :
Db       1579 ----TLRRDNGGQDRLV 1591


RESULT 14
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-33

  Query Match             42.1%;  Score 681.5;  DB 9;  Length 3739;
  Best Local Similarity   48.9%;  Pred. No. 3.5e-48;
  Matches  152;  Conservative   42;  Mismatches   98;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||| |  |  ::  |||| |||||||:| |:| :| || |:|||
Db       2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| |||||||| |::
Db       2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188
```

```
Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:|: :   :: |:| :       |:||||||| :|||||    : :|
Db       2189 GAHLAGQGGMLSLALSEAAVVERLAGFD-----GLSVAAVNGPTATVVSGDPTQIQELAQ 2243

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||| :||  :  ||  :   |  ::|:  ::|| |:: |  :    |
Db       2244 ACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPAL 2303

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| | |          ||
Db       2304 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPE------TVTG---- 2353

Qy        300 VIMGTLRRGQG 310
               :|||||  |
Db       2354 --LGTLRRDNG 2362


RESULT 15
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             42.1%;  Score 681.5;  DB 9;  Length 3739;
  Best Local Similarity   48.9%;  Pred. No. 3.5e-48;
  Matches  152;  Conservative   42;  Mismatches   98;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              ||||||| || ||| |||| |  |  ::  |||| |||||||:| |:| :| || |:|||
Db       2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA- 120
              ||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||| ||||||| |::
Db       2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||| | ::|:|: :   :: |:| :       |:||||||| :|||||    : :|
Db       2189 GAHLAGQGGMLSLALSEAAVVERLAGFD-----GLSVAAVNGPTATVVSGDPTQIQELAQ 2243

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  || |:|||| :||  :  ||  :   |  ::|:  ::|| |:: |  :  .  |
Db       2244 ACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPAL 2303

Qy        241 DAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |  |||||:   | | ||   |   :|   |||:| |||||| | |          ||
Db       2304 DGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPE------TVTG---- 2353

Qy        300 VIMGTLRRGQG 310
               :|||||  |
Db       2354 --LGTLRRDNG 2362


Search completed: June 17, 2003, 13:46:39
Job time : 10.8459 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 5.74347 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_29_344
Perfect score: 1618
Sequence: 1 VFVFPGQGPQWPGMGRELLD.....ADTVIMGTLRRGQGTLDHFL 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	819	50.6	344	3	US-08-858-003-33	Sequence 33, Appl
2	819	50.6	344	3	US-09-078-166-33	Sequence 33, Appl
3	819	50.6	344	4	US-08-997-467-33	Sequence 33, Appl
4	809.5	50.0	3724	2	US-08-804-227C-10	Sequence 10, Appl
5	809.5	50.0	3724	2	US-08-804-198-4	Sequence 4, Appli
6	776.5	48.0	3567	2	US-07-642-734C-4	Sequence 4, Appli
7	776.5	48.0	3567	3	US-08-439-009A-4	Sequence 4, Appli
8	775.5	47.9	829	4	US-09-413-814-105	Sequence 105, App
9	765	47.3	2595	4	US-09-036-987A-2	Sequence 2, Appli
10	765	47.3	2595	4	US-09-370-700-2	Sequence 2, Appli
11	762	47.1	3729	2	US-08-804-227C-4	Sequence 4, Appli
12	750	46.4	5588	4	US-09-036-987A-6	Sequence 6, Appli
13	750	46.4	5588	4	US-09-370-700-6	Sequence 6, Appli
14	726	44.9	1580	2	US-08-804-227C-11	Sequence 11, Appl
15	726	44.9	1580	2	US-08-804-198-5	Sequence 5, Appli
16	722	44.6	3491	2	US-07-642-734C-2	Sequence 2, Appli
17	722	44.6	3491	3	US-08-439-009A-2	Sequence 2, Appli
18	714.5	44.2	3816	4	US-09-428-517-3	Sequence 3, Appli
19	712	44.0	4150	4	US-09-428-517-2	Sequence 2, Appli
20	707	43.7	3519	4	US-09-428-517-4	Sequence 4, Appli
21	702.5	43.4	3170	2	US-07-642-734C-5	Sequence 5, Appli
22	702.5	43.4	3170	3	US-08-439-009A-5	Sequence 5, Appli
23	696.5	43.0	4551	3	US-09-320-878-1	Sequence 1, Appli
24	696.5	43.0	4613	4	US-09-105-537-31	Sequence 31, Appl
25	696.5	43.0	11877	4	US-09-105-537-6	Sequence 6, Appli
26	681.5	42.1	3739	3	US-09-320-878-2	Sequence 2, Appli
27	681.5	42.1	3739	4	US-09-105-537-33	Sequence 33, Appl
28	674.5	41.7	3170	4	US-09-036-987A-4	Sequence 4, Appli
29	674.5	41.7	3170	4	US-09-370-700-4	Sequence 4, Appli
30	674	41.7	1562	3	US-09-320-878-3	Sequence 3, Appli
31	674	41.7	1562	4	US-09-105-537-35	Sequence 35, Appl
32	669.5	41.4	6095	4	US-09-144-085-2	Sequence 2, Appli
33	667.5	41.3	1346	3	US-09-320-878-4	Sequence 4, Appli
34	667.5	41.3	1346	4	US-09-105-537-37	Sequence 37, Appl
35	667	41.2	345	3	US-08-858-003-34	Sequence 34, Appl
36	667	41.2	345	3	US-09-078-166-34	Sequence 34, Appl
37	667	41.2	345	4	US-08-997-467-34	Sequence 34, Appl
38	662	40.9	5087	4	US-09-144-085-1	Sequence 1, Appli
39	652.5	40.3	841	4	US-09-413-814-107	Sequence 107, App
40	627.5	38.8	2890	4	US-09-413-814-67	Sequence 67, Appl
41	627.5	38.8	3798	3	US-09-335-409-6	Sequence 6, Appli
42	627.5	38.8	3798	4	US-09-568-102-6	Sequence 6, Appli
43	627.5	38.8	3798	4	US-09-567-969-6	Sequence 6, Appli
44	627.5	38.8	3798	4	US-09-568-480-6	Sequence 6, Appli
45	627.5	38.8	3798	4	US-09-568-486-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-08-858-003-33
; Sequence 33, Application US/08858003
; Patent No. 6060234
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,003
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6060234e
US-08-858-003-33
```

```
  Query Match             50.6%;  Score 819;  DB 3;  Length 344;
  Best Local Similarity   54.0%;  Pred. No. 4.9e-73;
  Matches  168;  Conservative   38;  Mismatches   95;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  ||    ||: |  || :  :|:||  ||: |||  |||  |||| ||||
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||  :|:||||:||: |||| ||:||| ||||||||:|  |||  |:||:|  |  |:|
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
               |||  | :|:|    :|| ||:  | :    ||||||| | |  |:|  ||:|:|||
Db        125 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 180

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
               ||    |  | || ||   ||| : |:   ::  | |: |||: |||:|:|||| ||  |
Db        181 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTE 240

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
               ||| ||||||   | || | | |:  |    |:| ||||:| : |::   | |      |
Db        241 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 295

Qy        300 VIMGTLRRGQG 310
               ::|||||  |
Db        296 AVLGTLRRRHG 306


RESULT 2
US-09-078-166-33
; Sequence 33, Application US/09078166
; Patent No. 6063561
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,166
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6063561e
US-09-078-166-33

  Query Match             50.6%;  Score 819;  DB 3;  Length 344;
  Best Local Similarity   54.0%;  Pred. No. 4.9e-73;
  Matches  168;  Conservative   38;  Mismatches   95;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  ||    ||: |  || :  :|:||  ||: |||  |||  |||| ||||
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||  :|:||||:||: |||| ||:||| ||||||||:|  |||  |:||:|  |  |:|
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
               |||  | :|:|    :|| ||:  | :    ||||||| | |  |:|  ||:|:|||
Db        125 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 180

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
               ||    |  | || ||   ||| : |:   ::  | |: |||: |||:|:|||| ||  |
Db        181 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTE 240

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
               ||| ||||||   | || | | |:  |    |:| ||||:| : |::   | |      |
Db        241 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 295

Qy        300 VIMGTLRRGQG 310
               ::|||||  |
Db        296 AVLGTLRRRHG 306


RESULT 3
US-08-997-467-33
; Sequence 33, Application US/08997467
; Patent No. 6200813
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,467
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:  16-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6200813e
US-08-997-467-33

  Query Match             50.6%;  Score 819;  DB 4;  Length 344;
  Best Local Similarity   54.0%;  Pred. No. 4.9e-73;
  Matches  168;  Conservative   38;  Mismatches   95;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  ||    ||: |  ||  :  :|:||   ||: |||   |||  |||| ||||
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||  :|:||||:||: |||| ||:||| ||||||||:|  |||  |:||:|  |  |:|
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 ||| | :|:|    :|| ||:  | :    |||||||| | |  |:|  ||:|:|||
Db        125 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 180

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
                ||    |  | || ||   ||| : |:   ::  | |: |||: |||:|:|||| ||  |
Db        181 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFFYSTVTGGLLDGTE 240

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
                ||| ||||||    | || | | |:  |   |:| ||||:| : |::   | |      |
Db        241 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 295

Qy        300 VIMGTLRRGQG 310
                ::|||||  |
Db        296 AVLGTLRRRHG 306


RESULT 4
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-10

  Query Match             50.0%;  Score 809.5;  DB 2;  Length 3724;
  Best Local Similarity   54.1%;  Pred. No. 1.5e-70;
  Matches  170;  Conservative   37;  Mismatches   94;  Indels   13;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELL---DASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGL 57
              ||||||||  |||  |    ||      |  | |: |||: |  |:: |||  |||  | || |
Db       2089 VFVFPGQGSQWPEMADGLLARSSGSGSFLETARACDLALRPHLGWSVLDVLRREPGAPSL 2148

Qy         58 DRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLW 117
              |||||||  || :|:|||  ||| |||| ||:||| ||||||:|:|  |:|  ||||:| |
Db       2149 DRVDVVQPVLFTMMVSLAETWRSLGVEPAAVVGHSQGEIAAAYVAGALTLDDAARIVALR 2208

Qy        118 SQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVAD 177
              |||   ||| | :|:|  :  :| ||: ||::    |||||||||| :  |||  :|:|:
Db       2209 SQAWLRLAGKGGMVAVTLSERDLRPRLEPWSD----RLAVAAVNGPETCAVSGDPDALAE 2264

Qy        178 LVADLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD 236
              |||:| |  |  | ||:||   ||| :  :|  :   | |: || | |||:|:|||| :|
Db       2265 LVAELGAEGVHARPIPGVDTAGHSPQVDTLEAHLRKVLAPVAPRTSDIPFYSTVTGGLID 2324

Qy        237 TRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGT 296
              | |||| ||||||   | || | | |:  |   |:| ||||:| : |||   | |
Db       2325 TAELDADYWYRNMREPVEFEQATRALIADGHDVFLESSPHPMLAVSLQETISDAG----- 2379

Qy        297 ADTVIMGTLRRGQG 310
                :   ::||||||||
Db       2380 SPAAVLGTLRRGQG 2393


RESULT 5
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  4:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-4

  Query Match             50.0%;  Score 809.5;  DB 2;  Length 3724;
  Best Local Similarity   54.1%;  Pred. No. 1.5e-70;
  Matches  170;  Conservative   37;  Mismatches   94;  Indels   13;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELL---DASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGL 57
              ||||||| ||| |   ||     |  | |: |||: |  |:: |||  |||  | || |
Db       2089 VFVFPGQGSQWPEMADGLLARSSGSGSFLETARACDLALRPHLGWSVLDVLRREPGAPSL 2148

Qy         58 DRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLW 117
              |||||||| || :|:|||  ||| |||| ||:||| ||||||:|:| |:| ||||:| | |
Db       2149 DRVDVVQPVLFTMMVSLAETWRSLGVEPAAVVGHSQGEIAAAYVAGALTLDDAARIVALR 2208

Qy        118 SQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVAD 177
              |||   ||| | :|:|  :  :| ||: ||::    |||||||||| :  |||  :|:|:
Db       2209 SQAWLRLAGKGGMVAVTLSERDLRPRLEPWSD----RLAVAAVNGPETCAVSGDPDALAE 2264

Qy        178 LVADLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD 236
              |||:| |  |  | || ||   ||| :  :|  :   | |: || | |||:|:|||| :|
Db       2265 LVAELGAEGVHARPIPGVDTAGHSPQVDTLEAHLRKVLAPVAPRTSDIPFYSTVTGGLID 2324

Qy        237 TRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGT 296
              | |||| ||||||   | || | | |:  |   |:| ||||:| : |||   | |
Db       2325 TAELDADYWYRNMREPVEFEQATRALIADGHDVFLESSPHPMLAVSLQETISDAG----- 2379

Qy        297 ADTVIMGTLRRGQG 310
              :   ::||||||||
Db       2380 SPAAVLGTLRRGQG 2393


RESULT 6
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3567 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-642-734C-4

  Query Match             48.0%;  Score 776.5;  DB 2;  Length 3567;
  Best Local Similarity   50.0%;  Pred. No. 2.7e-67;
  Matches  158;  Conservative   44;  Mismatches  107;  Indels    7;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||:||||| || ||| ||| :| ||   :|||: : ||  || |  ||| :| |||||||
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:||| |||| |||| ||:||| ||||||||:| |:| |||::|   |:
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                :|:| | : :||      :  |: || :    ||:|||||||| ||||   |:
Db        680 MRSLSGEGGMAAVALGEAAVRERLRPWQD----RLSVAAVNGPRSVVVSGEPGALRAFSE 735

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              |  | :| | | ||  :||| :  : | ::      | |||:|: |||:|    :|  ||
Db        736 DCAAEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTEL 795

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:  ||||  |   | : |   |:|:|||||:   ::|    : :  |  | |
Db        796 DARYWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEA---VEEADGAEDAV 852

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|:| |  | |  ||
Db        853 VVGSLHRDGGDLSAFL 868


RESULT 7
US-08-439-009A-4
; Sequence 4, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3567 amino acids
;      TYPE:  amino acid
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-439-009A-4

  Query Match             48.0%;  Score 776.5;  DB 3;  Length 3567;
  Best Local Similarity   50.0%;  Pred. No. 2.7e-67;
  Matches  158;  Conservative   44;  Mismatches  107;  Indels    7;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||:||||| || ||| ||| :| ||   :|||: : ||  || |  ||| :| |||||||
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||||:||| |||| |||| ||:||| ||||||||:| |:| |||::|   |:
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              :|:| | : :||     :  |: || :    ||:|||||||| ||||   |:
Db        680 MRSLSGEGGMAAVALGEAAVRERLRPWQD----RLSVAAVNGPRSVVVSGEPGALRAFSE 735

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              | | :| | | ||  :||| :  : | ::    | |||:|: |||:|    :|  ||
Db        736 DCAAEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTEL 795

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:  |||| |   | : |   |:|:|||||:   ::|    : :  |  | |
Db        796 DARYWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEA---VEEADGAEDAV 852

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|:| |  | |  ||
Db        853 VVGSLHRDGGDLSAFL 868


RESULT 8
US-09-413-814-105
; Sequence 105, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 105
;   LENGTH: 829
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-105

  Query Match             47.9%;  Score 775.5;  DB 4;  Length 829;
  Best Local Similarity   52.8%;  Pred. No. 3.7e-68;
  Matches  168;  Conservative   46;  Mismatches   91;  Indels   13;  Gaps    4;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVL--RDSPDAPGLDR 59
              |||||| ||| | | || :|  ||  : ||| | ||:||||:  ||   ::  || |:|
Db        338 FVFPGQGSQWPSMARALLASSPAFRAEIEACERALAPHVDWSLLAVLGGDEAHAAPMLER 397

Qy         60 VDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQ 119
              |||||| ||||||:||| ||: || | ||:||| ||||||:|:| ||| |||||| | |:
Db        398 VDVVQPVLFAVMIALAATWRAAGVTPDAVVGHSQGEIAAAYVAGALSLEDAARVVALRSR 457

Qy        120 AQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
              | | ||| ||: :|  |  ||  |:||  |    |||:||:| | : :|||   |: :||
Db        458 AITKLAGRGAMSAVELTTAELEARLAPLGE----RLAIAAINSPHAALVSGDPGAIDELV 513

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              |:|: ||:  | : |:  :||  : |||  ::  |  | |||: :| :|:||| |||
Db        514 AELSGAQLFARKVRVEYASHSAHVEAIERTMLEALDGIAPRPATVPLYSAVTGERLDGEA 573

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDL-GDTTGTAD 298
              | ||:||||:  ||||| | | ||  | : |||:|||||||: | |    : || |
Db        574 LGAAHWYRNLRHTVRFEHATRALLDDGHRFFVEVSPHPVLTVALDETLSSIDGDAT---- 629

Qy        299 TVIMGTLRRGQGTLDHFL 316
              :: :||| :| |: ||
Db        630 --VVSSLRRDEGDLERFL 645


RESULT 9
US-09-036-987A-2
; Sequence 2, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2595 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-2

  Query Match             47.3%;  Score 765;  DB 4;  Length 2595;
  Best Local Similarity   50.6%;  Pred. No. 2.3e-66;
  Matches  157;  Conservative   51;  Mismatches   92;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||||:||: |:||  |:  |  | ||::|||:  |||::  |  ||
```

```
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||:|:||| |||| || | ||:||| |||||| |:| ||: |||||| : |:
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              | |||:||: |:    :|:   : || :    |: || |||| ||:||| |||:|:|:|
Db        647 LTALAGSGAMASLQHPAEEVRQILLPWRD----RIGVAGVNGPSSTLVSGDREAMAELLA 702

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :   ::| | |||:  :||| :  : : ::  | |: ||   || :|: ||  || | :
Db        703 ECADRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPM 761

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:   | || |   ||::|   |:|:|||||||  :|| |   |       : |
Db        762 DADYWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAG-----REVV 816

Qy        301 IMGTLRRGQG 310
              :||||||:|
Db        817 ALGTLRRGEG 826


RESULT 10
US-09-370-700-2
; Sequence 2, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2595
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-2

  Query Match             47.3%;  Score 765;  DB 4;  Length 2595;
  Best Local Similarity   50.6%;  Pred. No. 2.3e-66;
  Matches  157;  Conservative   51;  Mismatches   92;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||||:||: |:||   |:   |  | ||::|||:  |||::   |  ||
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||:|:||| |||| || | ||:||| |||||| |:| ||: |||||| : |:
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              | |||:||: |:    :|:   : || :    |: || |||| ||:||| |||:|:|:|
Db        647 LTALAGSGAMASLQHPAEEVRQILLPWRD----RIGVAGVNGPSSTLVSGDREAMAELLA 702

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :   ::| | |||:  :||| :  : : ::  | |: ||   || :|: ||  || | :
Db        703 ECADRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPM 761

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:   | || |   ||::|   |:|:|||||||  :|| |   |       : |
Db        762 DADYWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAG-----REVV 816

Qy        301 IMGTLRRGQG 310
              :||||||:|
Db        817 ALGTLRRGEG 826


RESULT 11
US-08-804-227C-4
; Sequence 4, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3729 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-4

  Query Match             47.1%;  Score 762;  DB 2;  Length 3729;
  Best Local Similarity   50.9%;  Pred. No. 7.8e-66;
  Matches  161;  Conservative   46;  Mismatches   99;  Indels   10;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| ||| | |:||| :  |||: :||:|| : ::|||| ||:: |||| | ||
Db       2096 VFVFPGQGSQWPSMARDLLDRAPAFRETAKACDAALSVHLDWSVLDVLQEKPDAPPLSRV 2155

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| || :|:|||| ||  || | ||:||| |||||| |:| ||| |||:| | |:|
Db       2156 DVVQPVLFTMMLSLAACWRDLGVHPAAVVGHSQGEIAAACVAGALSLEDAARIVALRSRA 2215

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              |||| | : :|:      |  ||  : :    ||:||||| | :  |:|  :|: :|:|
Db       2216 WLTLAGKGGMAAVSLPEARLRERIERFGQ----RLSVAAVNSPGTAAVAGDVDALRELLA 2271

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              :|||  :| : || ||   ||  :  ::| :   | |::|| | |||:|:|||  |||
Db       2272 ELTAEGIRAKPIPGVDTAGHSAQVDGLKEHLFEVLAPVSPRSSDIPFYSTVTGAPLDTER 2331

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||| ||||||   | || | | |:  |   |:| :|||:| | | |   | |   ||
Db       2332 LDAGYWYRNMREPVEFEKAVRALIADGYDLFLECNPHPMLAMSLDETLTDSGG-HGT--- 2387

Qy        300 VIMGTLRRGQGTLDHF 315
              :| |||| :|:   |
Db       2388 -VMHTLRRQKGSAKDF 2402
```

```
RESULT 12
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match             46.4%;  Score 750;  DB 4;  Length 5588;
  Best Local Similarity   50.2%;  Pred. No. 2.2e-64;
  Matches  156;  Conservative   45;  Mismatches   96;  Indels   14;  Gaps    3;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              |||||||  | || ||| ::| |||  :  | || |||||||:  |||  |||| ||| |
Db        563 FVFPGQGSHWQGMARELSESSPVFRRKLAECAAATAPYVDWSLLGVLRGDPDAPALDRDD 622

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQ 121
              |:|  |||:|:||| |||| |||| ||:||| ||||||||:| ||| || |::     |
Db        623 VIQLALFAMMVSLAELWRSCGVEPAAVVGHSQGEIAAAHVAGALSLTDAVRIIAARCDAV 682

Qy        122 TTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVAD 181
              : | | | ::::|     :: |||    |      | ||||||| ||||||   |:  |  :
Db        683 SALTGKGGMLAIALPESAVVKRIAGLPE-----LTVAAVNGPGSTVVSGEPSALERLQTE 737

Qy        182 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELD 241
              |||  |:|| : :|  :||| :  :: |::  |  :   |: | |:|:||| | ||  ||
Db        738 LTAENVQTRRVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAEIAFYSTVTGERTDTGRLD 797

Qy        242 AAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT-- 299
              | |||:|:   |||:     :  || : |||:||||:|| |:||       |   ||
Db        798 ADYWYQNLRQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGIQE-------TLEAADAGG 850

Qy        300 VIMGTLRRGQG 310
              |::|:||||:|
Db        851 VVVGSLRRGEG 861


RESULT 13
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match             46.4%;  Score 750;  DB 4;  Length 5588;
  Best Local Similarity   50.2%;  Pred. No. 2.2e-64;
  Matches  156;  Conservative   45;  Mismatches   96;  Indels   14;  Gaps    3;

Qy          2 FVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVD 61
              |||||||  | || ||| ::| |||  :  | || |||||||:  |||  |||| ||| |
Db        563 FVFPGQGSHWQGMARELSESSPVFRRKLAECAAATAPYVDWSLLGVLRGDPDAPALDRDD 622

Qy         62 VVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQ 121
              |:|  |||:|:||| |||| |||| ||:||| ||||||||:| ||| || |::     |
Db        623 VIQLALFAMMVSLAELWRSCGVEPAAVVGHSQGEIAAAHVAGALSLTDAVRIIAARCDAV 682

Qy        122 TTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVAD 181
              : | | | ::::|     :: |||    |      | ||||||| ||||||   |:  |  :
Db        683 SALTGKGGMLAIALPESAVVKRIAGLPE-----LTVAAVNGPGSTVVSGEPSALERLQTE 737

Qy        182 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELD 241
              |||  |:|| : :|  :||| :  :: |::  |  :   |: | |:|:||| | ||  ||
Db        738 LTAENVQTRRVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAEIAFYSTVTGERTDTGRLD 797

Qy        242 AAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT-- 299
              | |||:|:   |||:     :  || : |||:||||:|| |:||       |   ||
Db        798 ADYWYQNLRQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGIQE-------TLEAADAGG 850

Qy        300 VIMGTLRRGQG 310
              |::|:||||:|
Db        851 VVVGSLRRGEG 861


RESULT 14
US-08-804-227C-11
; Sequence 11, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
```

```
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1580 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-11

  Query Match             44.9%;  Score 726;  DB 2;  Length 1580;
  Best Local Similarity   48.7%;  Pred. No. 7.9e-63;
  Matches  154;  Conservative   37;  Mismatches  115;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  |||  ||||  :  |||      |   |  |  ::||||  ||||:|  ||  :||
Db        565 VFVFPGQGSQWTGMGAELLDRAPVFRAKAEECARALAAHLDWSVLDVLRDAPGAPPIDRA 624

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||||||  :|:||||||  |  ||  |  ||:|||  |||||||  :|  |||  |||||:    |:
Db        625 DVVQPTLFTMMVSLAALWESHGVRPAAVVGHSQGEIAAAHAAGALSLDDAARVIAERSRL 684

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 |||  |  ::||  |   |  :    :  ||  |     |::|||||||  |   |:|    |: :
Db        685 WKRLAGNGGMLSVMAPADRVRELMEPWAE----RMSVAAVNGPASVTVAGDARALEEFGG 740

Qy        181 DLTAAQV-RTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
                |:||  |  |   :   ||   |||  :       ::   |  :|:|| |:||     :
Db        741 RLSAAGVLRWPLAGVDFAGHSPQVEQFRAELLDTLGTVRPTAARLPFFSTVTAAAHEPEG 800

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||||||||||    |  |     | ||::|  :|||||  |||:|    :  |:|    |
Db        801 LDAAYWYRNMREPVEFASTLRTLLREGHRTFVEMGPHPLLGAAIDEVAEAEG-----VHA 855

Qy        300 VIMGTLRRGQGTLDHF 315
                :  ||  ||  |  ||  |
Db        856 TALATLHRGSGGLDRF 871


RESULT 15
US-08-804-198-5
; Sequence 5, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1580 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-5

  Query Match             44.9%;  Score 726;  DB 2;  Length 1580;
  Best Local Similarity   48.7%;  Pred. No. 7.9e-63;
  Matches  154;  Conservative   37;  Mismatches  115;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||  |||  ||||  :  |||      |   |  |  ::||||  ||||:|  ||  :||
Db        565 VFVFPGQGSQWTGMGAELLDRAPVFRAKAEECARALAAHLDWSVLDVLRDAPGAPPIDRA 624

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||||||  :|:||||||  |  ||  |  ||:|||  |||||||  :|  |||  |||||:    |:
Db        625 DVVQPTLFTMMVSLAALWESHGVRPAAVVGHSQGEIAAAHAAGALSLDDAARVIAERSRL 684

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 |||  |  ::||  |   |  :    :  ||  |     |::|||||||  |   |:|    |: :
Db        685 WKRLAGNGGMLSVMAPADRVRELMEPWAE----RMSVAAVNGPASVTVAGDARALEEFGG 740

Qy        181 DLTAAQV-RTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
                |:||  |  |   :   ||   |||  :       ::   |  :|:|| |:||     :
Db        741 RLSAAGVLRWPLAGVDFAGHSPQVEQFRAELLDTLGTVRPTAARLPFFSTVTAAAHEPEG 800

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||||||||||    |  |     | ||::|  :|||||  |||:|    :  |:|    |
Db        801 LDAAYWYRNMREPVEFASTLRTLLREGHRTFVEMGPHPLLGAAIDEVAEAEG-----VHA 855

Qy        300 VIMGTLRRGQGTLDHF 315
                :  ||  ||  |  ||  |
Db        856 TALATLHRGSGGLDRF 871


Search completed: June 17, 2003, 13:09:24
Job time : 6.74347 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 15.48 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_29_344
Perfect score: 1618
Sequence: 1 VFVFPGQGPQWPGMGRELLD.....ADTVIMGTLRRGQGTLDHFL 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1618	100.0	3972	21	AAB23749	S. avermitilis ave
2	1618	100.0	3972	22	AAG65264	Streptomyces averm
3	1618	100.0	3972	22	AAG65268	Streptomyces averm
4	819	50.6	344	20	AAW87714	An active acyltran
5	819	50.6	344	21	AAB12926	Protein sequence o
6	809.5	50.0	3724	18	AAW23718	Platenolide syntha
7	809.5	50.0	3724	18	AAW22608	Platenolide syntha
8	791	48.9	4881	21	AAB23751	S. avermitilis ave
9	791	48.9	4881	22	AAG65266	Streptomyces averm
10	776.5	48.0	3567	14	AAR44431	eryA region polype
11	765.5	47.3	3192	22	AAE10128	Streptomyces nours
12	765	47.3	2595	20	AAY39297	SpnA a polyketide
13	765	47.3	2595	22	AAB70965	S. spinosa protein
14	762	47.1	3729	18	AAW22603	Tylactone synthase
15	753	46.5	7068	22	AAE10142	Streptomyces nours
16	753	46.5	9477	22	AAE10144	Streptomyces nours
17	750	46.4	5588	20	AAY39301	SpnE a polyketide
18	750	46.4	5588	22	AAB70969	S. spinosa protein
19	746	46.1	3562	22	AAB82213	Polyketide synthas
20	734	45.4	5532	21	AAB23752	S. avermitilis ave
21	734	45.4	5532	22	AAG65267	Streptomyces averm
22	726	44.9	1580	18	AAW23719	Platenolide syntha
23	726	44.9	1580	18	AAW22609	Platenolide syntha
24	718.5	44.4	3201	22	AAB82214	Polyketide synthas
25	714.5	44.2	3816	21	AAY92708	S. antibioticus 8,
26	713	44.1	6239	21	AAB23750	S. avermitilis ave
27	713	44.1	6239	22	AAG65265	Streptomyces averm
28	712	44.0	4150	21	AAY92707	S. antibioticus 8,
29	707	43.7	3519	21	AAY92709	S. antibioticus 8,
30	704	43.5	3546	22	AAB82212	Polyketide synthas
31	702.5	43.4	2986	14	AAR44432	eryA region polype
32	698	43.1	3398	14	AAR44430	eryA region polype
33	697.5	43.1	5069	19	AAW52846	A. mediterranei ri
34	696.5	43.0	4551	21	AAB18637	Amino acid sequenc
35	696.5	43.0	4551	21	AAY67201	Narbonolide syntha
36	696.5	43.0	4613	21	AAY77192	S. venezuelae macr
37	696.5	43.0	4613	21	AAY77200	S. venezuelae pik
38	696.5	43.0	12199	21	AAY77180	S. venezuelae pik
39	682	42.2	383	23	AAU77689	Actinomycete monen
40	681.5	42.1	3739	21	AAB18638	Amino acid sequenc
41	681.5	42.1	3739	21	AAY77193	S. venezuelae macr
42	681.5	42.1	3739	21	AAY77201	S. venezuelae pik
43	681.5	42.1	3739	21	AAY67202	Narbonolide syntha
44	674.5	41.7	3170	20	AAY39299	SpnC a polyketide
45	674.5	41.7	3170	22	AAB70967	S. spinosa protein

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 1618; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 3.2e-138;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
Db 29 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 88

Qy 61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
Db 89 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 148

Qy 121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
Db 149 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 208

Qy 181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
Db 209 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 268

Qy 241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
Db 269 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 328

Qy 301 IMGTLRRGQGTLDHFL 316
Db 329 IMGTLRRGQGTLDHFL 344

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 1618; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 3.2e-138;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
Db 29 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 88

Qy 61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
Db 89 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 148

Qy 121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
Db 149 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 208

Qy 181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
Db 209 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 268

Qy 241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
Db 269 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 328

Qy 301 IMGTLRRGQGTLDHFL 316
Db 329 IMGTLRRGQGTLDHFL 344

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.

```
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79279.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 5; Page 235-248; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is derived from an S.
CC   avermitilis protein.
XX
SQ   Sequence   3972 AA;

  Query Match             100.0%;  Score 1618;  DB 22;  Length 3972;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-138;
  Matches  316;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         29 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 88

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 148

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 208

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 268

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 328

Qy        301 IMGTLRRGQGTLDHFL 316
              ||||||||||||||||
Db        329 IMGTLRRGQGTLDHFL 344


RESULT 4
AAW87714
ID   AAW87714 standard; Protein; 344 AA.
XX
AC   AAW87714;
XX
DT   09-MAR-1999  (first entry)
XX
DE   An active acyltransferase domain (NidAT5) from a polyketide synthase.
XX
KW   Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW   polyketide synthesis; antimicrobial; anticancer; antifungal;
KW   immunosuppressant; anthelmintic; NidAT5; ethyl acyltransferase domain;
KW   module 5; PKS cluster.
XX
OS   Streptomyces caelestis.
XX
PN   WO9851695-A2.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-US09518.
XX
PR   16-MAY-1997;   97US-0858003.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Kakavas SJ,  Katz L,  Pereda-Lopez A,  Ruan X,  Stassi DL;
PI   Summers RG;
XX
DR   WPI; 1999-070114/06.
DR   N-PSDB; AAV99255.
XX
PT   New polyketides produced by microorganism having altered polyketide
PT   synthase genes - especially chimeras containing a heterologous
PT   acyltransferase domain, useful as e.g. antimicrobials, anticancer
PT   agents
XX
PS   Claim 25; Fig 21; 137pp; English.
XX
CC   The present sequence represents an enzymatically active acyltransferase
CC   domain, NidAT5, from a polyketide synthase (PKS). The NidAT5 protein
CC   is derived from the ethyl acyltransferase domain from module 5 of the
CC   PKS cluster of Streptomyces caelestis NRRL-2821. The sequence
CC   can be used to replace acyltransferase domains of other PKS enzymes,
CC   which are then used to generate novel polyketides by de novo biosynthesis
CC   rather than by chemical modification. Polyketides of the invention have
CC   the methyl groups on the macrolactone ring substituted with -H, -Et
CC   and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC   dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC   antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC   anthelmintic agents, or their intermediates.
XX
SQ   Sequence   344 AA;

  Query Match             50.6%;  Score 819;  DB 20;  Length 344;
  Best Local Similarity   54.0%;  Pred. No. 1.5e-66;
  Matches  168;  Conservative   38;  Mismatches   95;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||    ||: |  || :  :|:||  ||: |||  |||  |||| ||||
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| || :|:||||:||: |||| ||:||| |||||||:| ||| |:||:| | |:|
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                ||| | :|:|    :|| ||:  | :    |||||||| | |  |:|  ||:|:|||
Db        125 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 180

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
               ||   |  | || ||   ||| : |:   ::  | |: |||: |||:|:|||| ||  |
Db        181 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTE 240

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||| ||||||   | || | | |:   |   |:| ||||:| : |::   | |      |
Db        241 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 295

Qy        300 VIMGTLRRGQG 310
               ::|||||  |
Db        296 AVLGTLRRRHG 306


RESULT 5
AAB12926
ID   AAB12926 standard; Protein; 344 AA.
XX
AC   AAB12926;
XX
```

DT 20-NOV-2000 (first entry)
XX
DE Protein sequence of the ethyl AT module of the niddamycin PKS cluster.
XX
KW Erythromycin derivative; polyketide synthase; PKS; acetyltransferase;
KW antibiotic; antifungal; anticancer; immunosuppressant; antihelminithic;
KW ethyl AT; niddamycin.
XX
OS Streptomyces caelestis.
XX
PN US6063561-A.
XX
PD 16-MAY-2000.
XX
PF 13-MAY-1998; 98US-0078166.
XX
PR 17-JAN-1991; 91US-0642734.
PR 16-MAY-1997; 97US-0858003.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Katz L, Stassi DL, Ruan X, Pereda-Lopez A, Kakavas SJ, Summers RG;
XX
DR WPI; 2000-386760/33.
DR N-PSDB; AAA72521.
XX
PT New erythromycin derivatives produced by microorganisms with altered
PT polyketide synthase genes -
XX
PS Example 20; Fig 21; 94pp; English.
XX
CC This invention relates to novel erythromycin derivatives, in which the
CC methyl groups on the macrolactone ring have been substituted with H, Et
CC and/or OH and the ethyl side chain has been substituted with
CC hydroxymethyl or dihydroxycyclohexylmethyl side chains. Included in the
CC invention are vectors containing polynucleotide sequences which encode an
CC enzymatically active acetyltransferase domain from Streptomyces. Also
CC included are methods for altering the substrate specificity of a
CC polyketide synthase in a microorganism using the vectors of the
CC invention. Polyketides are are a large class of natural products
CC including antibiotics, antifungal, anticancer and immunosuppressant
CC compounds. Polyketide synthase (PKS) is a complex of enzyme activities
CC responsible for the biosynthesis of polyketides. The erythromycin
CC derivatives exhibit antibiotic, antifungal, anticancer, antihelminithic
CC and immunosuppressant activity.
CC The present sequence represents the amino acid sequence of the ethyl
CC acetyltransferase (AT) module of the niddamycin PKS cluster isolated from
CC Streptomyces caelestis. The sequence is used in the production of a
CC vector of the invention.
XX
SQ Sequence 344 AA;

Query Match 50.6%; Score 819; DB 21; Length 344;
Best Local Similarity 54.0%; Pred. No. 1.5e-66;
Matches 168; Conservative 38; Mismatches 95; Indels 10; Gaps 3;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||   ||: |  || :  :|:||  ||: |||  |||  |||| ||||
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| || :|:||||:||: |||| ||:||| |||||||||:| ||| |:||:| | |:|
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                 ||| | :|:|    :|| ||:  | :    |||||||| | |  |:|  ||:|:|||
Db        125 WLGLAGKGGMVAVPMPAEELRPRLVTWGD----RLAVAAVNSPGSCAVAGDPEALAELVA 180

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
               ||   |  | |:| ||   ||| : |:   ::  | |: |||: |||:|:|:|||| ||  |
Db        181 LLTGEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTE 240

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              |||| ||||||   | || | | |:  |   |:| ||||:| : |::   | |      |
Db        241 LDATYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAG-----TDA 295

Qy        300 VIMGTLRRGQG 310
              ::|||||  |
Db        296 AVLGTLRRRHG 306
```

RESULT 6
AAW23718
ID AAW23718 standard; Protein; 3724 AA.
XX
AC AAW23718;
XX
DT 27-FEB-1998 (first entry)
XX
DE Platenolide synthase ORF3 protein.
XX
KW Platenolide synthase gene cluster; platenolide production; srmG gene;
KW multi-functional protein; macrolide antibiotic; spiramycin.
XX
OS Streptomyces ambofaciens.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "encoded by TGT"
FT Domain 36..459
FT /note= "ketosynthase domain, KS4"
FT Domain 594..921
FT /note= "acyltransferase domain, AT4"
FT Domain 1177..1350
FT /note= "ketosynthase domain, KS04"
FT Domain 1459..1523
FT /note= "acyl carrier protein domain, ACP4"
FT Domain 1545..1969
FT /note= "ketosynthase domain, KS5"
FT Domain 2088..2424
FT /note= "acyltransferase domain, AT5"
FT Domain 2445..2619
FT /note= "dehydratase domain, DH5"
FT Domain 2912..3261
FT /note= "enoylreductase domain, ER5"
FT Domain 3266..3451
FT /note= "ketoreductase domain, KR5"
FT Domain 3551..3631
FT /note= "acyl carrier protein domain, ACP5"
XX
PN EP791656-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-0301066.
XX
PR 22-FEB-1996; 96US-0012050.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI Rosteck PR;
XX
DR WPI; 1997-418047/39.
DR N-PSDB; AAT78508.
XX
PT DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT for production of spiramycin-related polyketide antibiotics
XX
PS Claim 8; Pages 54-66; 81pp; English.
XX
CC AAW23716-W23720 represent proteins encoded by the platenolide synthase
CC gene cluster of the invention. The gene cluster is also referred to as
CC the srmG gene, and was isolated from Streptomyces ambofaciens. These
CC sequences are multi-functional proteins which direct the synthesis of
CC the polyketide platenolide. Platenolide is the basic building block of

CC the macrolide antibiotic spiramycin. The DNA can be used to produce
CC compounds exhibiting antibiotic activity based on the platenolide
CC structure, including specifically the macrolide antibiotic spiramycin and
CC spiramycin analogues and derivatives. Modifications of the platenolide
CC synthase DNA sequence can be made so as to change the number and type of
CC carboxylic acids incorporated into the growing polyketide chain and to
CC change the kind of post-condensation processing that is conducted.
XX
SQ Sequence 3724 AA;

Query Match 50.0%; Score 809.5; DB 18; Length 3724;
Best Local Similarity 54.1%; Pred. No. 2.4e-64;
Matches 170; Conservative 37; Mismatches 94; Indels 13; Gaps 4;

Qy 1 VFVFPGQGPQWPGMGRELL---DASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGL 57
Db 2089 VFVFPGQGSQWPEMADGLLARSSGSGSFLETARACDLALRPHLGWSVLDVLRREPGAPSL 2148

Qy 58 DRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLW 117
Db 2149 DRVDVVQPVLFTMMVSLAETWRSLGVEPAAVVGHSQGEIAAAYVAGALTLDDAARIVALR 2208

Qy 118 SQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVAD 177
Db 2209 SQAWLRLAGKGGMVAVTLSERDLRPRLEPWSD----RLAVAAVNGPETCAVSGDPDALAE 2264

Qy 178 LVADLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD 236
Db 2265 LVAELGAEGVHARPIPGVDTAGHSPQVDTLEAHLRKVLAPVAPRTSDIPFYSTVTGGLID 2324

Qy 237 TRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGT 296
Db 2325 TAELDADYWYRNMREPVEFEQATRALIADGHDVFLESSPHPMLAVSLQETISDAG----- 2379

Qy 297 ADTVIMGTLRRGQG 310
Db 2380 SPAAVLGTLRRGQG 2393

RESULT 7
AAW22608
ID AAW22608 standard; Protein; 3724 AA.
XX
AC AAW22608;
XX
DT 27-FEB-1998 (first entry)
XX
DE Platenolide synthase ORF3 protein.
XX
KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW polyketide; tylactone synthesis; antibiotic; tylosin.
XX
OS Streptomyces ambofaciens.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "encoded by TGT"
FT Domain 36..459
FT /note= "ketosynthase domain, KS4"
FT Domain 594..921
FT /note= "acyltransferase domain, AT4"
FT Domain 1177..1350
FT /note= "ketosynthase domain, KS04"
FT Domain 1459..1523
FT /note= "acyl carrier protein domain, ACP4"
FT Domain 1545..1969
FT /note= "ketosynthase domain, KS5"
FT Domain 2088..2424
FT /note= "acyltransferase domain, AT5"
FT Domain 2445..2619
FT /note= "dehydratase domain, DH5"
FT Domain 2912..3261
FT /note= "enoylreductase domain, ER5"
FT Domain 3266..3451
FT /note= "ketoreductase domain, KR5"
FT Domain 3551..3631
FT /note= "acyl carrier protein domain, ACP5"
XX
PN EP791655-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-0301056.
XX
PR 22-FEB-1996; 96US-0012078.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX
DR WPI; 1997-418046/39.
DR N-PSDB; AAT80414.
XX
PT DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
PS Example 2; Pages 155-167; 220pp; English.
XX
CC AAW22606-W22610 represent proteins encoded by the platenolide synthase
CC gene cluster. The gene cluster is also referred to as the srmG gene, and
CC was isolated from Streptomyces ambofaciens. These sequences are
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The DNA encoding this sequence was used along
CC with the tylG gene (see AAT80413) to create a hybrid ORF1 sequence (see
CC AAT80415). The tylG gene is the tylactone synthase gene cluster of the
CC invention. The tylG sequence was isolated from Streptomyces fradiae, and
CC encodes multifunctional proteins which direct the synthesis of the
CC polyketide tylactone. Tylactone is the basic building block of the
CC antibiotic tylosin. The hybrid sequence can be used to transform S.
CC ambofaciens lacking the srmG ORF1 sequence, or S. fradiae lacking the
CC tylG ORF1 sequence, so that they can produce polyketides. The DNA
CC sequence can be modified so as to alter the type of carboxylic acids
CC incorporated, the number of carboxylic acids incorporated and/or the
CC post-condensation reactions performed, thereby resulting in novel
CC tylosin-related polyketides.
XX
SQ Sequence 3724 AA;

Query Match 50.0%; Score 809.5; DB 18; Length 3724;
Best Local Similarity 54.1%; Pred. No. 2.4e-64;
Matches 170; Conservative 37; Mismatches 94; Indels 13; Gaps 4;

Qy 1 VFVFPGQGPQWPGMGRELL---DASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGL 57
Db 2089 VFVFPGQGSQWPEMADGLLARSSGSGSFLETARACDLALRPHLGWSVLDVLRREPGAPSL 2148

Qy 58 DRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLW 117
Db 2149 DRVDVVQPVLFTMMVSLAETWRSLGVEPAAVVGHSQGEIAAAYVAGALTLDDAARIVALR 2208

Qy 118 SQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVAD 177
Db 2209 SQAWLRLAGKGGMVAVTLSERDLRPRLEPWSD----RLAVAAVNGPETCAVSGDPDALAE 2264

Qy 178 LVADLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD 236
Db 2265 LVAELGAEGVHARPIPGVDTAGHSPQVDTLEAHLRKVLAPVAPRTSDIPFYSTVTGGLID 2324

Qy 237 TRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGT 296
Db 2325 TAELDADYWYRNMREPVEFEQATRALIADGHDVFLESSPHPMLAVSLQETISDAG----- 2379

Qy 297 ADTVIMGTLRRGQG 310
Db 2380 SPAAVLGTLRRGQG 2393

```
RESULT 8
AAB23751
ID   AAB23751 standard; Protein; 4881 AA.
XX
AC   AAB23751;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:5.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 254-278; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   4881 AA;

  Query Match             48.9%;  Score 791;  DB 21;  Length 4881;
  Best Local Similarity   52.2%;  Pred. No. 1.7e-62;
  Matches  164;  Conservative   45;  Mismatches   97;  Indels    8;  Gaps    2;

Qy          3 VPPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDV 62
              :| |||  | || :||| :| || : :  |  |  ||:|||:  ||| :|||| | ||||
Db       2139 LFGGQGTHWEGMAQELLGSSPVFAQQMSDCAQALEPYLDWSLLDVLRGAPDAPPLQRVDV 2198

Qy         63 VQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQT 122
              ||| |||||:|||||||| || | || ||| |||||:|:| ||| |||||  | |||
Db       2199 VQPVLFAVMVSLAALWRSYGVHPDAVAGHSQGEIAAAYVAGALSLDDAARVTALRSQALA 2258

Qy        123 TLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADL 182
               ||| ||: ||    ::| ||:| | :    || :||||| || ||||  |||  || :|
Db       2259 ALAGQGAMASVGLPVEKLEPRLATWGD----RLVIAAVNGARSAVVSGEPEAVDALVEEL 2314

Qy        183 TAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDA 242
              :   |  | : ||  :||| : ||: |::  | ||  |   :||:|:||| |:|  ||||
Db       2315 SHEDVPARRLMVDWASHSPQVEAIQGRLLELLAPIRARTGDVPFYSTVTGERIDGTELDA 2374

Qy        243 AYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIM 302
               |||||:   |||  | : |:: |   |:|  ||| : :|:||   ::||     |:::
Db       2375 DYWYRNLRQVVRFRDATQALVRAGHTVFIEACPHPAVAVGVQETLDEMGD----LDSLVV 2430

Qy        303 GTLRRGQGTLDHFL 316
               |:||||:| |  ||
Db       2431 GSLRRGEGGLRRFL 2444


RESULT 9
AAG65266
ID   AAG65266 standard; Protein; 4881 AA.
XX
AC   AAG65266;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 6.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 201-217; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   4881 AA;

  Query Match             48.9%;  Score 791;  DB 22;  Length 4881;
  Best Local Similarity   52.2%;  Pred. No. 1.7e-62;
  Matches  164;  Conservative   45;  Mismatches   97;  Indels    8;  Gaps    2;

Qy          3 VPPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDV 62
              :| |||  | || :||| :| || : :  |  |  ||:|||:  ||| :|||| | ||||
Db       2139 LFGGQGTHWEGMAQELLGSSPVFAQQMSDCAQALEPYLDWSLLDVLRGAPDAPPLQRVDV 2198

Qy         63 VQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQT 122
              ||| |||||:|||||||| || | || ||| |||||:|:| ||| |||||  | |||
Db       2199 VQPVLFAVMVSLAALWRSYGVHPDAVAGHSQGEIAAAYVAGALSLDDAARVTALRSQALA 2258

Qy        123 TLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADL 182
               ||| ||: ||    ::| ||:| | :    || :||||| || ||||  |||  || :|
Db       2259 ALAGQGAMASVGLPVEKLEPRLATWGD----RLVIAAVNGARSAVVSGEPEAVDALVEEL 2314
```

```
Qy        183 TAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDA 242
Db       2315 SHEDVPARRLMVDWASHSPQVEAIQGRLLELLAPIRARTGDVPFYSTVTGERIDGTELDA 2374

Qy        243 AYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIM 302
Db       2375 DYWYRNLRQVVRFRDATQALVRAGHTVFIEACPHPAVAVGVQETLDEMGD----LDSLVV 2430

Qy        303 GTLRRGQGTLDHFL 316
Db       2431 GSLRRGEGGLRRFL 2444
```

RESULT 10
AAR44431
ID AAR44431 standard; Protein; 3567 AA.
XX
AC AAR44431;
XX
DT 22-DEC-1993 (first entry)
XX
DE eryA region polypeptide module #2.
XX
KW Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW erythromycin; condensation; elongation; acyl chain growth;
KW gene replacement.
XX
OS Saccarapolyspora erythraea.
XX
PN WO9313663-A.
XX
PD 22-JUL-1993.
XX
PF 17-JAN-1992; 92WO-US00427.
XX
PR 17-JAN-1992; 92WO-US00427.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Donadio S, Katz L, McAlpine JB;
XX
DR WPI; 1993-242804/30.
DR N-PSDB; AAQ46806.
XX
PT Biosynthesis of specific polyketide analogues esp. erythromycin
PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT microorganisms
XX
PS Disclosure; Fig 2; 133pp; English.
XX
CC The sequences given in AAR44430-32 are encoded by the eryA fragment of
CC the Saccarapolyspora erythraea genome. These polypeptides are
CC involved in the biosynthesis of the polyketide segment of erythromycin.
CC eryA is organised in modules and each module takes care of one
CC condensation step. The precise succession of elongation steps is
CC dictated by the genetic order of the modules. The DNA encoding
CC these polypeptides may be specifically altered such that novel
CC polyketide molecules of desired structure are produced. Three types
CC of alteration may be produced; those inactivating a single function in
CC a module which does not arrest acyl chain growth; those inactivating a
CC single function in a module which does affect chain growth; and those
CC affecting an entire module. The mutations may be introduced by gene
CC replacement.
XX
SQ Sequence 3567 AA;

Query Match 48.0%; Score 776.5; DB 14; Length 3567;
Best Local Similarity 50.0%; Pred. No. 2.3e-61;
Matches 158; Conservative 44; Mismatches 107; Indels 7; Gaps 2;

```
Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
Db        680 MRSLSGEGGMAAVALGEAAVRERLRPWQD----RLSVAAVNGPRSVVVSGEPGALRAFSE 735

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
Db        736 DCAAEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTEL 795

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
Db        796 DARYWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEA---VEEADGAEDAV 852

Qy        301 IMGTLRRGQGTLDHFL 316
Db        853 VVGSLHRDGGDLSAFL 868
```

RESULT 11
AAE10128
ID AAE10128 standard; Protein; 3192 AA.
XX
AC AAE10128;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin gene, NysB.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; PKS type I.
XX
OS Streptomyces noursei.
XX

FH	Key	Location/Qualifiers
FT	Domain	42..462
FT		/label= KS1_domain
FT		/note= "Ketosynthase (KS) domain"
FT	Domain	578..897
FT		/label= AT1_domain
FT		/note= "Acyltransferase (AT) domain"
FT	Domain	911..1110
FT		/label= DH1_domain
FT		/note= "Dehydratase (DH) domain"
FT	Domain	1201..1447
FT		/label= KR1_domain
FT		/note= "Ketoreductase (KR) domain"
FT	Domain	1484..1559
FT		/label= ACP1_domain
FT		/note= "Acyl carrier protein (ACP) domain"
FT	Domain	1579..2004
FT		/label= KS2_domain
FT		/note= "Ketosynthase (KS) domain"
FT	Domain	2117..2439
FT		/label= AT2_domain
FT		/note= "Acyltransferase (AT) domain"
FT	Domain	2453..2659
FT		/label= DH2_domain
FT		/note= "Dehydratase (DH) domain"
FT	Domain	2749..2996
FT		/label= KR2_domain
FT		/note= "Ketoreductase (KR) domain"
FT	Domain	3025..3102
FT		/label= ACP2_domain
FT		/note= "Acyl carrier protein (ACP) domain"

XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX

```
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17184.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 168-169; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin
CC   gene, NysB.
XX
SQ   Sequence   3192 AA;

  Query Match             47.3%;  Score 765.5;  DB 22;  Length 3192;
  Best Local Similarity   50.6%;  Pred. No. 2e-60;
  Matches  160;  Conservative   42;  Mismatches  109;  Indels    5;  Gaps    2;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  || |||||||  |:||| ::  |  | |||||||:  ||    |   |:||
Db       2114 VFVFPGQGSQWWGMGRELLATSEVFRTAIDDCATALAPYVDWSLHDVLAGEGDPALLERV 2173

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  |||:|: |:||||| || | ||:||| |||||| |:| ||||||||||  | |||
Db       2174 DVVQPALFAMMVGLSALWRSHGVVPAAVVGHSQGEIAAACVAGALSLADAARVVALRSQA 2233

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
                |:| | ::||:|  : :   :||| |      |:||||||| | ||||  :|:  |
Db       2234 LPQLSGRGGMMSVSAPVERVTALLAPWQE----ALSVAAVNGPSSVVVSGDTDALDALHT 2289

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
                  || | :||  :|   : |: : :   | |: ||   :||:|:||| |:|
Db       2290 ACQEQGVRARKVSVDYASHGRHVEAVRDELARVLAPVDPRAPEVPFYSTVTGDRVDDAAF 2349

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              | |||| |:  ||| | | | ||  | : |:|:|||||   :||    : : ||    |
Db       2350 DGAYWYTNLRQTVRMEEATRALLAAGHRVFIEVSPHPVLAAPIQETQEAVAEATG-GSAV 2408

Qy        301 IMGTLRRGQGTLDHFL 316
              ::|:||| :|    ||
Db       2409 VLGSLRRDEGGPRRFL 2424


RESULT 12
AAY39297
ID   AAY39297 standard; Protein; 2595 AA.
XX
AC   AAY39297;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnA a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1944
FT                   /note= "Encoded by CAG"
FT   Domain          6..423
FT                   /label= KSi
FT                   /note= "Beta-ketosynthase domain: part of the
FT                         initiator module"
FT   Domain          528..853
FT                   /label= ATi
FT                   /note= "Acyl transferase domain: part of the
FT                         initiator module"
FT   Domain          895..977
FT                   /label= ACPi
FT                   /note= "Acyl carrier protein domain: part of the
FT                         initiator module"
FT   Domain          998..1413
FT                   /label= KS1
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 1"
FT   Domain          1525..1858
FT                   /label= AT1
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 1"
FT   Domain          2158..2337
FT                   /label= KR1
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 1"
FT   Domain          2432..2513
FT                   /label= ACP1
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 1"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 87-96; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnA gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnA gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
```

```
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   2595 AA;

  Query Match             47.3%;  Score 765;  DB 20;  Length 2595;
  Best Local Similarity   50.6%;  Pred. No. 1.7e-60;
  Matches  157;  Conservative   51;  Mismatches   92;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||||:||: |:|| |:  |  | ||::|||:  |||::  |  ||
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||:|:||| |||| || | ||:||| |||||| |:| ||: |||||| : |:
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              | |||:||: |:    :|:   : || :    |: || |||| ||:||| |||:|:|:|
Db        647 LTALAGSGAMASLQHPAEEVRQILLPWRD----RIGVAGVNGPSSTLVSGDREAMAELLA 702

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :   ::| | |||:  :||| :  : : ::  | |: ||   || :|: ||  || | :
Db        703 ECADRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPM 761

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:   | || |   ||::|   |:|:|||||| :|| |   |      : |
Db        762 DADYWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAG-----REVV 816

Qy        301 IMGTLRRGQG 310
               :||||||:|
Db        817 ALGTLRRGEG 826


RESULT 13
AAB70965
ID   AAB70965 standard; Protein; 2595 AA.
XX
AC   AAB70965;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF18, SEQ ID 41.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88335.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 57; Page 179-189; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   2595 AA;

  Query Match             47.3%;  Score 765;  DB 22;  Length 2595;
  Best Local Similarity   50.6%;  Pred. No. 1.7e-60;
  Matches  157;  Conservative   51;  Mismatches   92;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              |||||||| || ||||:||: |:|| |:  |  | ||::|||:  |||::  |  ||
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| |||:|:||| |||| || | ||:||| |||||| |:| ||: |||||| : |:
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              | |||:||: |:    :|:   : || :    |: || |||| ||:||| |||:|:|:|
Db        647 LTALAGSGAMASLQHPAEEVRQILLPWRD----RIGVAGVNGPSSTLVSGDREAMAELLA 702

Qy        181 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL 240
              :   ::| | |||:  :||| :  : : ::  | |: ||   || :|: ||  || | :
Db        703 ECADRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPM 761

Qy        241 DAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTV 300
              || |||||:   | || |   ||::|   |:|:|||||| :|| |   |      : |
Db        762 DADYWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAG-----REVV 816

Qy        301 IMGTLRRGQG 310
               :||||||:|
Db        817 ALGTLRRGEG 826


RESULT 14
AAW22603
ID   AAW22603 standard; Protein; 3729 AA.
```

```
XX
AC   AAW22603;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase ORF3 protein.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "encoded by GTG"
FT   Domain          43..465
FT                   /note= "ketosynthase domain, KS4"
FT   Domain          588..916
FT                   /note= "acyltransferase domain, AT4"
FT   Domain          1178..1356
FT                   /note= "ketoreductase domain, KR'4"
FT   Domain          1451..1534
FT                   /note= "acyl carrier protein domain, ACP4"
FT   Domain          1556..1980
FT                   /note= "ketosynthase domain, KS5"
FT   Domain          2095..2425
FT                   /note= "acyltransferase domain, AT5"
FT   Domain          2451..2636
FT                   /note= "dehydratase domain, DH2"
FT   Domain          3274..3453
FT                   /note= "ketoreductase domain, KR5"
FT   Domain          3548..3631
FT                   /note= "acyl carrier protein domain, ACP5"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80413.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 9; Pages 86-98; 220pp; English.
XX
CC   AAW22601-W22605 represent proteins encoded by the tylactone synthase
CC   gene cluster of the invention. The gene cluster is also referred to as
CC   the tylG gene, and was isolated from Streptomyces fradiae. These
CC   sequences are multifunctional proteins which direct the synthesis of the
CC   polyketide tylactone, isolated from Streptomyces fradiae. Tylactone is
CC   the basic building block of the antibiotic tylosin. The DNA sequence can
CC   be modified so as to alter the type of carboxylic acids incorporated,
CC   the number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence   3729 AA;

  Query Match             47.1%;  Score 762;  DB 18;  Length 3729;
  Best Local Similarity   50.9%;  Pred. No. 5.2e-60;
  Matches  161;  Conservative   46;  Mismatches   99;  Indels   10;  Gaps    4;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||||  ||| | |:||| :  |||: :||:|| : ::||||  ||:: |||| | ||

Db       2096 VFVFPGQGSQWPSMARDLLDRAPAFRETAKACDAALSVHLDWSVLDVLQEKPDAPPLSRV 2155

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              ||||| || :|:|||| ||  || | ||:||| |||||| |:| ||| ||||:| | |:|

Db       2156 DVVQPVLFTMMLSLAACWRDLGVHPAAVVGHSQGEIAAACVAGALSLEDAARIVALRSRA 2215

Qy        121 QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVA 180
              |||| | : :|:      |  ||  : :    ||:||||| | :  |:|  :|: :|:|

Db       2216 WLTLAGKGGMAAVSLPEARLRERIERFGQ----RLSVAAVNSPGTAAVAGDVDALRELLA 2271

Qy        181 DLTAAQVRTRMIP-VDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              :|||  :| : || ||   ||  :  ::| :   | |::|| | |||:|:|||  |||
Db       2272 ELTAEGIRAKPIPGVDTAGHSAQVDGLKEHLFEVLAPVSPRSSDIPFYSTVTGAPLDTER 2331

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              ||| ||||||   | || | | |:  |   |:| :|||:| | | |   | |   ||
Db       2332 LDAGYWYRNMREPVEFEKAVRALIADGYDLFLECNPHPMLAMSLDETLTDSGG-HGT--- 2387

Qy        300 VIMGTLRRGQGTLDHF 315
               :| |||| :|:   |
Db       2388 -VMHTLRRQKGSAKDF 2402


RESULT 15
AAE10142
ID   AAE10142 standard; Protein; 7068 AA.
XX
AC   AAE10142;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (partial).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          34..448
FT                   /label= KS9_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          572..890
FT                   /label= AT9_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          904..1123
FT                   /label= DH9_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1443..1686
FT                   /label= KR9_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1720..1793
FT                   /label= ACP9_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1813..2236
FT                   /label= KS10_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2346..2664
FT                   /label= AT10_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2678..2890
FT                   /label= DH10_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          2983..3229
FT                   /label= KR10_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3266..3339
FT                   /label= ACP10_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3358..3780
FT                   /label= KS11_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          3898..4217
```

```
FT                   /label= AT11_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4231..4432
FT                   /label= DH11_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4523..4770
FT                   /label= KR11_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          4806..4879
FT                   /label= ACP11_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          4801..5325
FT                   /label= KS12_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          5432..5754
FT                   /label= AT12_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          5768..5977
FT                   /label= DH12_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          6068..6315
FT                   /label= KR12_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          6348..6421
FT                   /label= ACP12_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          6454..6873
FT                   /label= KS13_domain
FT                   /note= "Ketosynthase (KS) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17185.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 181-185; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (partial).
XX
SQ   Sequence   7068 AA;

  Query Match           46.5%;  Score 753;  DB 22;  Length 7068;
  Best Local Similarity  53.0%;  Pred. No. 7.9e-59;

  Matches  168;  Conservative   34;  Mismatches  105;  Indels   10;  Gaps    3;

Qy          1 VFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRV 60
              ||||||| || ||| :||| | || | :  | || | : |||:  |||    || |:||
Db       3895 VFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAALAEFTDWSLVDVLRGVVGAPSLERV 3954

Qy         61 DVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQA 120
              |||||  ||||:||||||||:|| | ||:||| |||||| ||| ||| | |||| | |||
Db       3955 DVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQGEIAAAVVSGALSLRDGARVVALRSQA 4014

Qy        121 -QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLV 179
               ||| | ::||| : | | ||:  :      |::|||||||| ||:|  ||:  |
Db       4015 IGRALAGRGGMMSVALSVDVLEPRLVEF----EGRVSVAAVNGPRSVVVAGEPEALDALH 4070

Qy        180 ADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRE 239
              | |||  :| | | ||  :||  :  : | ::  | .: || | :|| |:|||  |||
Db       4071 ARLTADDIRARRIAVDYASHSHQVEDLHEELLEVLAELAPRTSEVPFFSTVTGDWLDTAR 4130

Qy        240 LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADT 299
              :|| ||:||:   |||  |   ||    : |||:| ||||:| :||   : |
Db       4131 MDAGYWFRNLRGRVRFADAVADLLAAEYRAFVEVSSHPVLSMAVQEAIDEAG-----VPA 4185

Qy        300 VIMGTLRRGQGTLDHFL 316
              |  ||||| ||  | ||
Db       4186 VAAGTLRRDQGGTDRFL 4202


Search completed: June 17, 2003, 13:02:27
Job time : 16.48 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 6.74779 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_2983_3128
Perfect score: 764
Sequence: 1 YWLESTQPGAGNVSAAGLDP.....ACGGGGASGGVWTCHASGVL 146

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	764	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	760	99.5	3626	2	Q9F7T9	Q9f7t9 streptomyce
3	585.5	76.6	5532	2	Q9S0R4	Q9s0r4 streptomyce
4	387.5	50.7	3576	16	Q9EX53	Q9ex53 streptomyce
5	386.5	50.6	3524	2	Q93H86	Q93h86 streptomyce
6	386	50.5	10917	2	Q93NW6	Q93nw6 streptomyce
7	385.5	50.5	3739	2	Q9ZGI4	Q9zgi4 streptomyce
8	381	49.9	9510	2	Q93NX9	Q93nx9 streptomyce
9	377	49.3	9507	2	Q9EWA1	Q9ewa1 streptomyce
10	372	48.7	9477	2	Q9L4X3	Q9l4x3 streptomyce
11	366.5	48.0	3970	2	Q93HI8	Q93hi8 streptomyce
12	365.5	47.8	5588	2	Q9ALM2	Q9alm2 saccharopol
13	362	47.4	6396	2	Q9KID7	Q9kid7 streptomyce
14	361.5	47.3	3591	2	Q9KIE1	Q9kie1 streptomyce
15	361	47.3	6420	2	P95814	P95814 streptomyce
16	357	46.7	4881	2	Q9S0R3	Q9s0r3 streptomyce
17	357	46.7	7525	2	Q9KIE0	Q9kie0 streptomyce
18	354.5	46.4	11096	2	Q9L4W3	Q9l4w3 streptomyce
19	351.5	46.0	6239	2	Q9S0R7	Q9s0r7 streptomyce
20	351	45.9	3939	2	Q93HJ3	Q93hj3 streptomyce
21	350	45.8	6048	2	Q93H87	Q93h87 streptomyce
22	348	45.5	6797	2	Q9X993	Q9x993 streptomyce
23	346.5	45.4	2103	16	Q9CD81	Q9cd81 mycobacteri
24	342	44.8	2152	2	Q9ALM5	Q9alm5 saccharopol
25	337.5	44.2	1907	2	Q93HI9	Q93hi9 streptomyce
26	336.5	44.0	6315	2	Q9ADL6	Q9adl6 polyangium
27	333.5	43.7	3613	2	Q93HJ1	Q93hj1 streptomyce
28	333.5	43.7	4472	2	O33954	O33954 streptomyce
29	332.5	43.5	1808	2	Q9EWA2	Q9ewa2 streptomyce
30	332	43.5	5069	2	O52789	O52789 amycolatops
31	329	43.1	5060	2	O52545	O52545 amycolatops
32	329	43.1	10223	2	Q54296	Q54296 streptomyce
33	324.5	42.5	4151	16	O53490	O53490 mycobacteri
34	322.5	42.2	2024	2	Q9EWA3	Q9ewa3 streptomyce
35	320	41.9	4613	2	Q9ZGI5	Q9zgi5 streptomyce
36	319	41.8	3816	2	Q9KIV3	Q9kiv3 streptomyce
37	317	41.5	7576	2	Q9ZGA4	Q9zga4 streptomyce
38	315.5	41.3	1937	2	O30482	O30482 streptomyce
39	314.5	41.2	8563	2	Q54297	Q54297 streptomyce
40	313.5	41.0	8817	2	Q53840	Q53840 polyangium
41	312	40.8	2126	16	P94996	P94996 mycobacteri
42	311.5	40.8	4928	2	Q9ALM3	Q9alm3 saccharopol
43	311.5	40.8	5644	2	Q93NX8	Q93nx8 streptomyce
44	309.5	40.5	5435	2	Q9L4X2	Q9l4x2 streptomyce
45	308.5	40.4	3413	2	O54593	O54593 amycolatops

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 764; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 9.2e-51;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2983 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 3042

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3043 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 3102

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |||||||||||||||||||||||||
Db       3103 YARGGSACGGGGASGGVWTCHASGVL 3128


RESULT 2
Q9F7T9
ID   Q9F7T9      PRELIMINARY;      PRT;  3626 AA.
AC   Q9F7T9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Avermectin polyketide synthase (Fragment).
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC31271;
RA   Hong Y.-S., Lee J.J.;
RT   "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT   and polyketide synthase gene from Streptomyces avermitilis.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF275943; AAG09812.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT   NON_TER    3626   3626
SQ   SEQUENCE   3626 AA;  380557 MW;  6272F5F088C1A8D0 CRC64;

  Query Match             99.5%;  Score 760;  DB 2;  Length 3626;
  Best Local Similarity   99.3%;  Pred. No. 1.7e-50;
  Matches  145;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2950 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 3009

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       3010 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGKGRRLVSV 3069

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |||||||||||||||||||||||||
Db       3070 YARGGSACGGGGASGGVWTCHASGVL 3095


RESULT 3
Q9S0R4
ID   Q9S0R4      PRELIMINARY;      PRT;  5532 AA.
AC   Q9S0R4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 3.
GN   AVEA3.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84478.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5532 AA;  575193 MW;  EC7B2069D44A4057 CRC64;

  Query Match             76.6%;  Score 585.5;  DB 2;  Length 5532;
  Best Local Similarity   76.0%;  Pred. No. 1e-36;
  Matches  111;  Conservative   15;  Mismatches   19;  Indels    1;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||::   ||:|:|||||:| ||||| ||::||   | || ||||||||||| |: |||
Db        955 YWLDA-PTGAGDVAAAGLEPAEHPLLAATVQLADTDGCLLTGRLSLRSHPWLGDYEVGGA 1013

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              |||||: |:|||:  |  ||| |:::||:|||||||| ||||||||||||||||||||
Db       1014 VLLSGSAFVELAVQVGERVGCTRIEQLTVHAPLVVPVGGGVSVQVGVAAADGEGRRLVSV 1073

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |||||||||||||||||||||||||
Db       1074 YARGGSACGGGGASGGVWTCHASGVL 1099


RESULT 4
Q9EX53
ID   Q9EX53      PRELIMINARY;      PRT;  3576 AA.
AC   Q9EX53;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative type I polyketide synthase.
GN   SCO6274 OR SC2C4.04C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
```

```
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL512902; CAC22145.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR003662; sub_transporter.
DR   InterPro; IPR001998; Xylose_isom.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   TIGRFAMs; TIGR00128; fabD; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR   PROSITE; PS00173; XYLOSE_ISOMERASE_2; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3576 AA;  377350 MW;  8A83C6F90AC9928B CRC64;

  Query Match             50.7%;  Score 387.5;  DB 16;  Length 3576;
  Best Local Similarity   54.7%;  Pred. No. 1.7e-21;
  Matches   81;  Conservative   15;  Mismatches   49;  Indels    3;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||: |:  | : :  || ||:||:|||||:||     |   |||||:||||||| | ||
Db        921 YWLDVTE-AAADAAGLGLTPTDHPILGATLDLADGEQTLFTSRLSLRTHPWLADHTVAGT 979

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               || |  |:|||  ||  :|| ||:|||| ||||:|   || :|: |  ||| |||:| |
Db        980 TLLPGTGFVELAALAGERLGCPRVEELTLSAPLVLPERDGVRIQLAVGEADGAGRRVVDV 1039

Qy        121 YAR--GGSACGGGGASGGVWTCHASGVL 146
              |||  ||    |  |    ||  | |||
Db       1040 YARPDGGDETPGAVAKARPWTALAKGVL 1067


RESULT 5
Q93H86
ID   Q93H86      PRELIMINARY;      PRT;  3524 AA.
AC   Q93H86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69304.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3524 AA;  366619 MW;  64A65759C461EBC4 CRC64;

  Query Match             50.6%;  Score 386.5;  DB 2;  Length 3524;
  Best Local Similarity   56.2%;  Pred. No. 1.9e-21;
  Matches   82;  Conservative   16;  Mismatches   39;  Indels    9;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||:  | ||:| |||||  :|| ||| :||   || || | |||:: |||||||| ||
Db       2611 YWLEAGAP-AGDVRAAGLDAADHPFLGAAVELPDAGGHLLTGLLSLQAQPWLADHAVLGT 2669

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               ||| |  |:||| ||    ||| |:|||| ||||:|  |||::|| | | |  ||| :||
Db       2670 VLLPGTAFVELAAHAAAETGCDLVEELTLQAPLVLPEHGGVALQVLVGAQDDSGRRPLSV 2729

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ::| |          | |  :|:|||
Db       2730 HSRAG--------DDGPWIRNATGVL 2747


RESULT 6
Q93NW6
ID   Q93NW6      PRELIMINARY;      PRT; 10917 AA.
AC   Q93NW6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphC.
GN   AMPHC.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73514.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00059; ADH_ZINC; UNKNOWN_1.
```

```
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
KW   Phosphopantetheine.
SQ   SEQUENCE   10917 AA;  1132905 MW;  15AC5956BB5810A1 CRC64;

  Query Match             50.5%;  Score 386;  DB 2;  Length 10917;
  Best Local Similarity   56.2%;  Pred. No. 7.3e-21;
  Matches   82;  Conservative   10;  Mismatches   48;  Indels    6;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              :|         |:| ||||   ||||||| :|||    | |  ||||| |||||||| | ||
Db       9958 FWPTGRAAAGGDVRAAGLGSAEHPLLGAAVELAAGEGVLFTGRLSLSSHPWLADHVVMGT 10017

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |    ||:|: ||  ||| ||:|||| ||||:|   | : ||| ||| |  ||| | |
Db      10018 VLLPGTALLEMAVRAGDEVGCGRVEELTLAAPLVLPERGALQVQVAVAAPDDSGRRTVGV 10077

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |:|        ||   ||  ||:|||
Db      10078 YSR------LEGADESVWAQHAAGVL 10097


RESULT 7
Q9ZGI4
ID   Q9ZGI4      PRELIMINARY;      PRT;  3739 AA.
AC   Q9ZGI4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAII.
GN   PIKAII.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR   EMBL; AF079138; AAC69330.1; -.
DR   HSSP; Q02054; 2AF8.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3739 AA;  387170 MW;  3D3910824DA5B080 CRC64;

  Query Match             50.5%;  Score 385.5;  DB 2;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 2.5e-21;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:|||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |         || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546


RESULT 8
Q93NX9
ID   Q93NX9      PRELIMINARY;      PRT;  9510 AA.
AC   Q93NX9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphI.
GN   AMPHI.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73501.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_6.
KW   Phosphopantetheine.
SQ   SEQUENCE   9510 AA;  992817 MW;  1A8E003A15D478CF CRC64;

  Query Match             49.9%;  Score 381;  DB 2;  Length 9510;
  Best Local Similarity   53.4%;  Pred. No. 1.5e-20;
  Matches   78;  Conservative   13;  Mismatches   49;  Indels    6;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||  : | || :: : ||    |||| | : ||   |||| ||:|| |||||||| | |
Db        907 YWPRAVQGGAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGA 966

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              || |  |||||: ||  |||||||||| ||||:|  ||| ||: :   |  ||| |:|
Db        967 TLLPGTAFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTV 1026

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              | |  :          || ||:|||
Db       1027 YGRPDA------DEDAPWTSHATGVL 1046


RESULT 9
Q9EWA1
ID   Q9EWA1      PRELIMINARY;      PRT;  9507 AA.
AC   Q9EWA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   PimS2 protein.
GN   PIMS2.
OS   Streptomyces natalensis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=68242;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20547809; PubMed=11094342;
RA   Aparicio J.F., Fouces R., Mendes M.V., Olivera N., Martin J.F.;
RT   "A complex multienzyme system encoded by five polyketide synthase
RT   genes is involved in the biosynthesis of the 26-membered polyene
RT   macrolide pimaricin in Streptomyces natalensis.";
RL   Chem. Biol. 7:895-905(2000).
DR   EMBL; AJ278573; CAC20921.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine.
SQ   SEQUENCE   9507 AA;  994228 MW;  57BD80C8AD37E89F CRC64;

  Query Match             49.3%;  Score 377;  DB 2;  Length 9507;
  Best Local Similarity   54.8%;  Pred. No. 3.2e-20;
  Matches   80;  Conservative   13;  Mismatches   47;  Indels    6;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :|     |:| ||||    |||| | : :|   | || |||| |:|||||||| ||
Db        913 YWPKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT 972

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |||||: ||   || ||:|||| ||||:| :||| ||| | : |  ||| |||
Db        973 VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAVSV 1032

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ::|       |     || ||:| |
Db       1033 HSR------PDGPEELPWTQHAAGTL 1052


RESULT 10
Q9L4X3
ID   Q9L4X3      PRELIMINARY;      PRT;  9477 AA.
AC   Q9L4X3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysI.
GN   NYSI.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71766.1; -.
DR   HSSP; P81989; 1QE5.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   9477 AA;  988132 MW;  0D2BCA5D7B265483 CRC64;

  Query Match             48.7%;  Score 372;  DB 2;  Length 9477;
  Best Local Similarity   51.4%;  Pred. No. 7.7e-20;
  Matches   75;  Conservative   18;  Mismatches   47;  Indels    6;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :: | |  :: : ||    |||| | : ||  || || |||| ::|||||| | ||
Db        914 YWPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRGT 973

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              || |  |||||: ||  ||||||:|||| |||::|  ||| ||: :   |  ||| |:|
Db        974 TLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRTVNV 1033

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :||  :           || ||:|||
Db       1034 HARPDT------GDDTPWTAHATGVL 1053


RESULT 11
Q93HI8
ID   Q93HI8      PRELIMINARY;      PRT;  3970 AA.
AC   Q93HI8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69199.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
```

```
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3970 AA;  412657 MW;  A93E583FAFA68C47 CRC64;

  Query Match             48.0%;  Score 366.5;  DB 2;  Length 3970;
  Best Local Similarity   53.4%;  Pred. No. 8e-20;
  Matches   78;  Conservative   14;  Mismatches   47;  Indels    7;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||    ||||:::|||| | |||||||  |   |  || |  ||||  | |||||||:| |
Db        947 YWLRGV-PGAGDMAAAGLAPAEHPLLGAVLVPAAGGGRLFTGRLSTDSLPWLADHAIGDT 1005

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |    |||| ||||:| | | :|||  |||:|  ||: ||: | | |  | |  ::
Db       1006 VLLPGTAVAELALWAGTYLGLDSVGDLTLELPLVLPSSGGLRVQLAVGAPDESGEREFAL 1065

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||:         ::  ||| || | |
Db       1066 YAQAEE------SAEEVWTRHAGGTL 1085


RESULT 12
Q9ALM2
ID   Q9ALM2      PRELIMINARY;      PRT;  5588 AA.
AC   Q9ALM2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase extender modules 8-10.
GN   SPNE.
OS   Saccharopolyspora spinosa.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23262.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5588 AA;  585629 MW;  C47F38F44BAC6A11 CRC64;

  Query Match             47.8%;  Score 365.5;  DB 2;  Length 5588;
  Best Local Similarity   47.9%;  Pred. No. 1.4e-19;
  Matches   70;  Conservative   21;  Mismatches   50;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||||  :  ||:: : ||   |||||| : ||  || || |:||::: |||||| |||
Db        906 YWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQPWLADHVVGGA 965

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              :|| |  |:|: : |   |||| ::||:|  |||:|  | | ||: |   |  ||| | |
Db        966 ILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVRV 1025

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :    :|          |||||:| |
Db       1026 H-----SCRDDAVPQDSWTCHATGTL 1046


RESULT 13
Q9KID7
ID   Q9KID7      PRELIMINARY;      PRT;  6396 AA.
AC   Q9KID7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FkbA.
GN   FKBA.
OS   Streptomyces hygroscopicus var. ascomyceticus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=132248;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20323220; PubMed=10863099;
RA   Wu K., Chung L., Revill W.P., Katz L., Reeves C.D.;
RT   "The FK520 gene cluster of streptomyces hygroscopicus var.
RT   ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual
RT   polyketide extender units.";
RL   Gene 251:81-90(2000).
DR   EMBL; AF235504; AAF86396.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000634; S/T_dehydrtse.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6396 AA;  669864 MW;  FF13BC2431D2E805 CRC64;

  Query Match             47.4%;  Score 362;  DB 2;  Length 6396;
  Best Local Similarity   51.4%;  Pred. No. 3e-19;
  Matches   75;  Conservative   15;  Mismatches   46;  Indels   10;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  |     :|: || :  |||||| : |   || :| ||||| ||||| :||| ||
Db        812 YWLRPT--SRADVTGAGQEQVAHPLLGAAVALPGTGGVVLTGRLSLASHPWLGEHAVDGT 869

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| || |||||  ||  |||| : || :  |||:|  |||:| | :|  |  ||| |:|
Db        870 VLLPGAAFLELAARAGDEVGCDLLHELVIETPLVLPATGGVAVSVEIAEPDDTGRRAVTV 929

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :||   :        |:|| || | |
Db        930 HARADGS--------GLWTRHAGGFL 947
```

```
RESULT 14
Q9KIE1
ID   Q9KIE1      PRELIMINARY;      PRT;  3591 AA.
AC   Q9KIE1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FkbC.
GN   FKBC.
OS   Streptomyces hygroscopicus var. ascomyceticus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=132248;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20323220; PubMed=10863099;
RA   Wu K., Chung L., Revill W.P., Katz L., Reeves C.D.;
RT   "The FK520 gene cluster of streptomyces hygroscopicus var.
RT   ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual
RT   polyketide extender units.";
RL   Gene 251:81-90(2000).
DR   EMBL; AF235504; AAF86392.1; -.
DR   HSSP; Q02054; 2AF8.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3591 AA;  373910 MW;  6CBF2D8C9CEBB3B0 CRC64;

  Query Match             47.3%;  Score 361.5;  DB 2;  Length 3591;
  Best Local Similarity   53.7%;  Pred. No. 1.8e-19;
  Matches   79;  Conservative   13;  Mismatches   44;  Indels   11;  Gaps    3;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLEL-ATDGGALLAGRLSLRSHPWLADHAVGG 59
              ||||:   || ::|||||    ||:| |   | | ||| :| ||:|||:|||||||||| |
Db       2465 YWLEAA--GATDLSAAGLTGAAHPMLAAVTALPADDGGVVLTGRISLRTHPWLADHAVRG 2522

Qy         60 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 119
              |||| |  |:|| : ||    ||  |||| : :||||||    | | | |   ||  ||| |:
Db       2523 TVLLPGTAFVELVIRAGDETGCGIVDELVIESPLVVPVTAAVDVSVTVEGADEAGRRPVT 2582

Qy        120 VYARGGSACGGGGASGGVWTCHASGVL 146
              |:||           | || |||| |
Db       2583 VHARTEGT--------GSWTRHASGTL 2601


RESULT 15
P95814
ID   P95814      PRELIMINARY;      PRT;  6420 AA.
AC   P95814;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   FK506 polyketide synthase.
GN   FKBA.
OS   Streptomyces SP.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1931;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA6548;
RX   MEDLINE=97217427; PubMed=9063448;
RA   Motamedi H., Cai S.J., Shafiee S.J., Elliston K.O.;
RT   "Structural organization of a multifunctional polyketide synthase
RT   involved in the biosynthesis of the macrolide immunosuppressant
RT   FK506.";
RL   Eur. J. Biochem. 244:74-80(1997).
DR   EMBL; Y10438; CAA71463.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001763; Rhodanese-like.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   SMART; SM00450; RHOD; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
SQ   SEQUENCE   6420 AA;  672222 MW;  25BCD2069456BEF1 CRC64;

  Query Match             47.3%;  Score 361;  DB 2;  Length 6420;
  Best Local Similarity   50.0%;  Pred. No. 3.6e-19;
  Matches   73;  Conservative   18;  Mismatches   45;  Indels   10;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||:|:|  || :::| ||  | ||:| ||  | |: | :| ||:|||:|||||||  :  |
Db       4309 YWIEAT--GAADLTALGLTDTAHPMLAATAALPTNDGTVLTGRISLRTHPWLADHTIQDT 4366

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |||| : ||  |||| :||| :  || :|| | | : | |  ||  | | |||
Db       4367 VLLPGTAFLELVMRAGDEVGCDTIDELVIETPLTLPVTGAVDLTVTVEQADATGHRPVSV 4426

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :||          |    || ||:| |
Db       4427 HAR--------PAGTDTWTRHATGTL 4444


Search completed: June 17, 2003, 13:07:57
Job time : 7.74779 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 1.59218 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_2983_3128
Perfect score: 764
Sequence: 1 YWLESTQPGAGNVSAAGLDP.....ACGGGGASGGVWTCHASGVL 146

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	246.5	32.3	3567	1	ERY2_SACER	Q03132 saccharopol
2	96.5	12.6	2110	1	MCAS_MYCBO	Q02251 mycobacteri
3	85	11.1	957	1	Y278_MYCTU	P56877 mycobacteri
4	85	11.1	1582	1	YU30_RALSO	Q8xv02 ralstonia s
5	85	11.1	3591	1	FHAB_BORPE	P12255 bordetella
6	84.5	11.1	389	1	A2AB_PROHA	O19054 procavia ca
7	84.5	11.1	1901	1	YZ08_MYCTU	O53553 mycobacteri
8	84	11.0	451	1	OCT6_RAT	P20267 rattus norv
9	83	10.9	449	1	OCT6_MOUSE	P21952 mus musculu
10	81.5	10.7	324	1	FABH_RHOCA	P30790 rhodobacter
11	80.5	10.5	498	1	YI18_MYCTU	Q50615 mycobacteri
12	80.5	10.5	914	1	WA22_MYCTU	O06794 mycobacteri
13	79	10.3	460	1	MB48_MYCTU	Q933k8 mycobacteri
14	77.5	10.1	226	1	DEOC_MYCLE	Q9cb45 mycobacteri
15	77.5	10.1	358	1	YM05_MYCTU	Q10394 mycobacteri
16	77.5	10.1	559	1	HUTU_PSEAE	Q9hu83 pseudomonas
17	77.5	10.1	904	1	DPO1_MYCTU	Q07700 mycobacteri
18	77.5	10.1	1286	1	AIDA_ECOLI	Q03155 escherichia
19	77.5	10.1	1876	1	PPSA_MYCTU	Q10977 mycobacteri
20	76.5	10.0	356	1	LPXD_RALSO	Q8xzi1 ralstonia s
21	76.5	10.0	1270	1	DDX9_HUMAN	Q08211 homo sapien
22	76	9.9	481	1	LORI_MOUSE	P18165 mus musculu
23	75.5	9.9	448	1	OCT6_HUMAN	Q03052 homo sapien
24	74	9.7	546	1	NU5M_CHLRE	P08739 chlamydomon
25	74	9.7	778	1	YQ34_MYCTU	P71933 mycobacteri
26	74	9.7	1084	1	HDA4_HUMAN	P56524 homo sapien
27	73.5	9.6	289	1	DAPF_MYCTU	O33231 mycobacteri
28	73.5	9.6	406	1	TRPB_RHIME	Q92tc9 rhizobium m
29	73.5	9.6	543	1	YP91_MYCTU	Q50630 mycobacteri
30	73.5	9.6	825	1	5E5_RAT	Q63003 rattus norv
31	73.5	9.6	2035	1	HMP2_YEREN	P48633 yersinia en
32	73	9.6	911	1	PERT_BORBR	Q03035 bordetella
33	72.5	9.5	142	1	PA2E_MOUSE	Q9qul3 mus musculu
34	72.5	9.5	151	1	DUT_YERPE	Q8zjp5 yersinia pe
35	72.5	9.5	388	1	ARGJ_CORGL	Q59280 corynebacte
36	72.5	9.5	426	1	HUTI_BACHD	Q9kbe4 bacillus ha
37	72.5	9.5	447	1	ODP2_RHIME	Q9r9n3 rhizobium m
38	72	9.4	436	1	BIOA_MYCLE	P45488 mycobacteri
39	72	9.4	473	1	YY33_MYCLE	P37391 mycobacteri
40	72	9.4	477	1	MAZ_HUMAN	P56270 homo sapien
41	71.5	9.4	330	1	HX11_HUMAN	P31314 homo sapien
42	71.5	9.4	397	1	PRLA_LYSEN	P00778 lysobacter
43	71.5	9.4	530	1	HLY_HAL17	P29143 halophilic
44	71	9.3	335	1	TRXB_MYCTU	P52214 mycobacteri
45	71	9.3	341	1	MRAW_RHIME	Q92nl4 rhizobium m

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             32.3%;  Score 246.5;  DB 1;  Length 3567;
  Best Local Similarity   39.0%;  Pred. No. 5e-13;
  Matches   57;  Conservative   24;  Mismatches   58;  Indels    7;  Gaps    3;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              :|    :|   :||| |:   |||||.| ::::   |||:  ||||    ||||:| |||
Db       2357 FWPRPHRP--ADVSALGVRGAEHPLLLAAVDVPGHGGAVFTGRLSTDEQPWLAEHVVGGR 2414

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               |: |:  ::||| ||  ||   ::|| |  |||:    |  ::: | | |  ||| : |
Db       2415 TLVPGSVLVDLALAAGEDVGLPVLEELVLQRPLVL-AGAGALLRMSVGAPDESGRRTIDV 2473

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
               :|    |      :   |: ||:| |
Db       2474 HAAEDVA----DLADAQWSQHATGTL 2495


RESULT 2
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             12.6%;  Score 96.5;  DB 1;  Length 2110;
  Best Local Similarity   24.3%;  Pred. No. 1.7;
  Matches   43;  Conservative   17;  Mismatches   52;  Indels   65;  Gaps    8;

Qy         21 TEHPLLGATLELATD----------GGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLE 70
              | |||||: : |  :           | ::|:         ||:|| |     | || : |
Db        894 TVHPLLGSHVRLTEEPERHVWQGDVGTSVLS---------WLSDHQVHNVAALPGAAYCE 944

Qy         71 LALHAGTYV--GCDRVDELTLHAPLVV----PVDGGVSVQV-GV------AAADGEGRRL 117
              :|| |   |       | ::|    |::    |:|   |:   ||         |||  |
Db        945 MALAAAAEVFGEAAEVRDITFEQMLLLDEQTPIDAVASIDAPGVVNFTVETNRDGETTRH 1004

Qy        118 VSVYARGGS------------------ACGG-------------GGASGGVWTCH 141
              :   |                      | |               | | ||: | |
Db       1005 ATAALRAAEDDCPPPGYDITALLQAHPHAVNGTAMRESFAERGVTLGAAFGGLTTAH 1061
```

```
RESULT 3
Y278_MYCTU
ID   Y278_MYCTU     STANDARD;      PRT;   957 AA.
AC   P56877;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv0278c precursor.
GN   RV0278C OR MT0291 OR MTV035.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021930; CAA17353.1; -.
DR   EMBL; AE006936; AAK44511.1; ALT_INIT.
DR   TIGR; MT0291; -.
DR   TubercuList; Rv0278c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    957       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV0278C.
FT   CONFLICT     40     40       M -> I (IN REF. 2).
FT   CONFLICT    158    163       MISSING (IN REF. 2).
FT   CONFLICT    807    807       R -> G (IN REF. 2).
SQ   SEQUENCE   957 AA;  81905 MW;  71EBABD417FBA47C CRC64;

  Query Match             11.1%;  Score 85;  DB 1;  Length 957;
  Best Local Similarity   24.1%;  Pred. No. 7.2;
  Matches   40;  Conservative   11;  Mismatches   55;  Indels   60;  Gaps    5;

Qy          4 ESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLL 63
              :|:  | |    ||        | ||  |  : | :|:||           |   ||
Db        301 DSSFGGVGGAGGAG------GLFGAGGEGGSGGHSLVAG----------GDGGAGGN--- 341

Qy         64 SGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYAR 123
              :|   |  |  ||   |                    |||     |:  | | |     ::
Db        342 AGMLALGAAGGAGGIGG----------------DGGTLTAGGIGGAGGAGGNAGLLFGS 384

Qy        124 GGS-----------------------ACGGGGASGGVWTCHASGV 145
              |||                       || | :|||    | |:
Db        385 GGSGGAGGFGFADGGQGGPGGNAGTVFGSGGAGGNGGVGQGFAGGI 430


RESULT 4
YU30_RALSO
ID   YU30_RALSO     STANDARD;      PRT;  1582 AA.
AC   Q8XV02;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein RSc3030 precursor.
GN   RSC3030 OR RS04727.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
CC   -!- SIMILARITY: BELONGS TO THE UPF0192 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL646073; CAD16739.1; -.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16   1582       HYPOTHETICAL PROTEIN RSC3030.
SQ   SEQUENCE   1582 AA;  170090 MW;  8683D148F5AE3C2A CRC64;

  Query Match             11.1%;  Score 85;  DB 1;  Length 1582;
  Best Local Similarity   22.9%;  Pred. No. 12;
  Matches   43;  Conservative   17;  Mismatches   50;  Indels   78;  Gaps    6;

Qy          4 ESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLL 63
              ||| | ||::    :|||  |:   |  |: : ||    | :| |:
Db        363 ESTAPAAGDIKLDVIDPTGAPVATTTRLSGETGA--DARFTLPSN-----------AQ 408

Qy         64 SGATFLELALHAGTYVGCDRVDE--------------------------LTLHAPLVVPV 97
              :|   |        || |  || |                        ::|  |   ||
Db        409 AGGYTLRFDYGGSTYGGAFRVAEYIKPHFDVNLSLDKAGYGTGEAIKGKISLRYPDGKPV 468

Qy         98 -DGGVSVQVGV-----------------------AAADGEG-------------RRLVS 119
               || ||| :                        ||:|              | :|:
Db        469 KDGKVSVSLRAQQVTMVEGELQYAGLFPVKLEQQELTTDGDGNAALTLPAAKEPSRYVVT 528

Qy        120 VYARGGSA 127
              |:|  |:|
Db        529 VFANDGAA 536


RESULT 5
```

```
FHAB_BORPE
ID   FHAB_BORPE     STANDARD;      PRT;  3591 AA.
AC   P12255;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Filamentous hemagglutinin.
GN   FHAB.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90355839; PubMed=2388559;
RA   Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT   "Genetic characterization of Bordetella pertussis filamentous
RT   haemagglutinin: a protein processed from an unusually large
RT   precursor.";
RL   Mol. Microbiol. 4:787-800(1990).
RN   [2]
RP   SEQUENCE OF 1-3261 FROM N.A.
RX   MEDLINE=89202384; PubMed=2539596;
RA   Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT   "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT   sequence and crucial role in adherence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC   -!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC       INFECTION.
CC   -!- SUBCELLULAR LOCATION: SURFACE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60351; AAA22974.1; -.
DR   EMBL; M60351; AAA22975.1; ALT_INIT.
DR   EMBL; M60351; AAA22976.1; ALT_INIT.
KW   Antigen; Hemagglutinin.
SQ   SEQUENCE   3591 AA;  367420 MW;  EF7418B30D6E5138 CRC64;

  Query Match             11.1%;  Score 85;  DB 1;  Length 3591;
  Best Local Similarity   30.7%;  Pred. No. 28;
  Matches   47;  Conservative   24;  Mismatches   54;  Indels   28;  Gaps    9;

Qy         11 GNVSA-AGLDPTEHPLLGATLELATDGGALLAGRLSLR--SHPWLADHAVG-------GT 60
              | ||| :|:       :||   | :||   ::|| ::|      ||| ::|       |
Db        792 GKVSAKSGIGLEGWGAVGAD-SLGSDGAISVSGRDAVRVDQARSLADISLGAEGGATLGA 850

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADG-----EGR 115
              | :|:    : :  |: | :     :|||   | | |  :|:|  | : |      ||
Db        851 VEAAGS----IDVRGGSTVAAN-----SLHANRDVRVSGKDAVRVTAATSGGGLHVSSGR 901

Qy        116 R--LVSVYARGGSACGGGASGGVWTCHASGVL 146
              :  | :| |||  |   |||  : :  ||| |
Db        902 QLDLGAVQARGALAL-DGGAGVALQSAKASGTL 933


RESULT 6
A2AB_PROHA
ID   A2AB_PROHA     STANDARD;      PRT;   389 AA.
AC   O19054;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor) (Fragment).
GN   ADRA2B.
OS   Procavia capensis habessinica (Abyssinian hyrax).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Hyracoidea; Procaviidae; Procavia.
OX   NCBI_TaxID=9814;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97357151; PubMed=9214502;
RA   Springer M.S., Cleven G.C., Madsen O.J., de Jong W.W., Waddell V.G.,
RA   Amrine H.M., Stanhope M.J.;
RT   "Endemic African mammals shake the phylogenetic tree.";
RL   Nature 388:61-64(1997).
CC   -!- FUNCTION: ALPHA-2 ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED INHIBITION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y12523; CAA73123.1; -.
DR   HSSP; P29274; 1MMH.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Multigene family;
KW   Phosphorylation; Lipoprotein; Palmitate.
FT   NON_TER       1      1
FT   TRANSMEM     <1     25       1 (POTENTIAL).
FT   DOMAIN       26     36       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     37     62       2 (POTENTIAL).
FT   DOMAIN       63     72       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     73     95       3 (POTENTIAL).
FT   DOMAIN       96    117       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    118    140       4 (POTENTIAL).
FT   DOMAIN      141    156       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    157    180       5 (POTENTIAL).
FT   DOMAIN      181    353       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    354    377       6 (POTENTIAL).
FT   DOMAIN      378    386       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    387   >389       7 (POTENTIAL).
FT   DOMAIN      281    292       ASP/GLU-RICH (ACIDIC).
FT   DISULFID     72    151       BY SIMILARITY.
FT   SITE         79     79       IMPLICATED IN LIGAND BINDING (BY
FT                                SIMILARITY).
FT   SITE        163    163       IMPLICATED IN CATECHOL AGONIST BINDING
FT                                (BY SIMILARITY).
FT   SITE        167    167       IMPLICATED IN CATECHOL AGONIST BINDING
FT                                (BY SIMILARITY).
FT   NON_TER     389    389
SQ   SEQUENCE   389 AA;  42528 MW;  4F2089EA140876E1 CRC64;

  Query Match             11.1%;  Score 84.5;  DB 1;  Length 389;
  Best Local Similarity   26.3%;  Pred. No. 3.2;
  Matches   44;  Conservative    8;  Mismatches   56;  Indels   59;  Gaps    5;

Qy          8 PGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA-----DHAVGGTVL 62
              | |       |      ||                | || |:| : | ||      |     |
Db        194 PRAKGAPGEGESKQPHP-------------LTAGPLALANPPTLATSLAVDGEANGHSK 239

Qy         63 LSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAA--------DGEG 114
              |:|    | :   ||           ||      |    | | : ||  |        : ||
Db        240 LTGEKERETSEDPGT---------PTLQRSWPVLPSSGQSQKKGVCGASPEEEAEGEEEG 290

Qy        115 RRLVSVYARGGSACG----------------------GGGASGGVW 138
              | :|| |   ||||                      | ||:|| |
Db        291 SRPLSVPASPASACGPHLQQPQGSQVLATLRGQVLLGRGVGAAGGQW 337
```

```
RESULT 7
YZ08_MYCTU
ID   YZ08_MYCTU     STANDARD;      PRT;  1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv3508 precursor.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL022022; CAA17745.1; -.
DR   TubercuList; Rv3508; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1901       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                              RV3508.
SQ   SEQUENCE   1901 AA;  147627 MW;  C7B1923D5D0146CD CRC64;

  Query Match             11.1%;  Score 84.5;  DB 1;  Length 1901;
  Best Local Similarity   31.8%;  Pred. No. 16;
  Matches   42;  Conservative    7;  Mismatches   60;  Indels   23;  Gaps    5;

Qy          7 QPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGA 66
              |  |||   :|| | |     |||       |||  ||           :  ||||    |
Db        508 QGGAGGAGSAGTDAT---ATGATGGTGFSGGAGGAGGAG-------GNTGVGGTNGSGGQ 557

Qy         67 TFLELALHAGTY--VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARG 124
                 |  ||    || |         |  :   ||   : |    | |:|    |  | |
Db        558 GGTGGAGGAGGAGGVGADN--------PTGIGGTGGTGGKGGAGGAGGQGG---SSGAGG 606

Qy        125 GSACGGGGASGG 136
               :   || | :||
Db        607 TNGSGGAGGTGG 618


RESULT 8
OCT6_RAT
ID   OCT6_RAT       STANDARD;      PRT;   451 AA.
AC   P20267;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Octamer-binding transcription factor 6 (OCT-6) (POU-domain
DE   transcription factor SCIP) (TST-1).
GN   POU3F1 OR OTF6 OR OTF-6 OR OCT-6 OR SCIP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90378306; PubMed=1975954;
RA   Monuki E.S., Kuhn R., Weinmaster G., Trapp B.D., Lemke G.;
RT   "Expression and activity of the POU transcription factor SCIP.";
RL   Science 249:1300-1303(1990).
RN   [2]
RP   SEQUENCE OF 265-369 FROM N.A.
RX   MEDLINE=89295573; PubMed=2739723;
RA   He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,
RA   Rosenfeld M.G.;
RT   "Expression of a large family of POU-domain regulatory genes in
RT   mammalian brain development.";
RL   Nature 340:35-42(1989).
RN   [3]
RP   REVISION TO 320.
RA   He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,
RA   Rosenfeld M.G.;
RL   Nature 340:662-662(1989).
RN   [4]
RP   SEQUENCE OF 35-451 FROM N.A.
RX   MEDLINE=91141528; PubMed=1705013;
RA   He X., Gerrero M.R., Simmons D.M., Park R.E., Lin C.R.,
RA   Swanson L.W., Rosenfeld M.G.;
RT   "Tst-1, a member of the POU domain gene family, binds the promoter of
RT   the gene encoding the cell surface adhesion molecule P0.";
RL   Mol. Cell. Biol. 11:1739-1744(1991).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE OCTAMER MOTIF
CC       ('ATTTGCAT'). THOUGHT TO BE INVOLVED IN EARLY EMBRYOGENESIS AND
CC       NEUROGENESIS. BOTH IN NORMAL DEVELOPMENT AND IN RESPONSE TO NERVE
CC       TRANSECTION, SCIP EXPRESSION WAS TRANSIENTLY ACTIVATED ONLY
CC       DURING THE PERIOD OF RAPID CELL DIVISION THAT SEPARATES THE
CC       PREMYELINATING AND MYELINATING PHASES OF SCHWANN CELL
CC       DIFFERENTIATION. IN COTRANSFECTION ASSAYS, SCIP ACTED AS A
CC       TRANSCRIPTIONAL REPRESSOR OF MYELIN SPECIFIC GENES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: NEURAL TISSUES AND TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.
CC       CLASS-3 SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M72711; AAA42118.1; -.
DR   EMBL; M63712; AAA42303.1; -.
DR   PIR; S05044; S05044.
DR   PIR; S05449; S05449.
DR   HSSP; P14859; 1OCT.
DR   TRANSFAC; T00969; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000327; POU_domain.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00157; pou; 1.
DR   PRINTS; PR00028; POUDOMAIN.
DR   ProDom; PD000010; Homeobox; 1.
DR   ProDom; PD000583; POU_domain; 1.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00352; POU; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
```

```
DR   PROSITE; PS00035; POU_1; 1.
DR   PROSITE; PS00465; POU_2; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation; Repressor;
KW   Homeobox.
FT   DOMAIN      251    321       POU.
FT   DNA_BIND    339    398       HOMEOBOX.
FT   CONFLICT    423    423       S -> G (IN REF. 4).
SQ   SEQUENCE   451 AA;  45496 MW;  E32626A759125EAA CRC64;

  Query Match             11.0%;  Score 84;  DB 1;  Length 451;
  Best Local Similarity   26.7%;  Pred. No. 4.1;
  Matches   47;  Conservative    9;  Mismatches   76;  Indels   44;  Gaps    8;

Qy          8 PGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRL-----SLRSHPWL---ADHAV-- 57
              || |   | |  |  ||   |    |   | ||        |  | ||   | | |
Db         12 PGGG---AGGTGPLMHPDAAAAAAAAAAAAERLHAGAAYREVQKLMHHEWLGAGAGHPVGL 68

Qy         58 ----------GGTVLLSGATFLELALHAGTYVGCDRVDE----LTLHAPLV--------- 94
                      ||    :|   ||      |   |  | |:       || ||
Db         69 AHPQWLPTGGGGGGDWAGGPHLEHGKAGGGSTG--RADDGGGGGGFHARLVHQGAAHAGA 126

Qy         95 VPVDGGVSVQVGVA------AADGEGRRLVSVYARGGSACGGGGASGGVWTCHASG 144
                || :  :| |      |  |   : : :||:     ||||   |:      |
Db        127 AWAQGGTAHHLGPAMSPSPGAGGGHQPQPLGLYAQAAYPGGGGGGLAGMLAAGGGG 182


RESULT 9
OCT6_MOUSE
ID   OCT6_MOUSE     STANDARD;      PRT;   449 AA.
AC   P21952;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Octamer-binding transcription factor 6 (OCT-6) (POU-domain
DE   transcription factor SCIP).
GN   POU3F1 OR OTF6 OR OTF-6 OR OCT6 OR SCIP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Brain;
RX   MEDLINE=91081324; PubMed=1979677;
RA   Meijer D., Graus A., Kraay R., Langeveld A., Mulder M.P., Grosveld G.;
RT   "The octamer binding factor Oct6: cDNA cloning and expression in
RT   early embryonic cells.";
RL   Nucleic Acids Res. 18:7357-7365(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=91006074; PubMed=1976514;
RA   Suzuki N., Rohdewohld H., Neuman T., Gruss P., Schoeler H.R.;
RT   "Oct-6: a POU transcription factor expressed in embryonal stem cells
RT   and in the developing brain.";
RL   EMBO J. 9:3723-3732(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91204458; PubMed=1840678;
RA   Zimmerman E.C., Jones C.M., Fet V., Hogan B.L.M., Magnuson M.A.;
RT   "Nucleotide sequence of mouse SCIP cDNA, a POU-domain transcription
RT   factor.";
RL   Nucleic Acids Res. 19:956-956(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92228768; PubMed=1565620;
RA   Hara Y., Rovescalli C., Kim Y., Nirenberg M.;
RT   "Structure and evolution of four POU domain genes expressed in mouse
RT   brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284(1992).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE OCTAMER MOTIF
CC       ('ATTTGCAT'). THOUGHT TO BE INVOLVED IN EARLY EMBRYOGENESIS AND
CC       NEUROGENESIS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN EMBRYONAL STEM CELLS AND IN THE
CC       DEVELOPING BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.
CC       CLASS-3 SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54628; CAA38445.1; -.
DR   EMBL; X56959; CAA40280.1; -.
DR   EMBL; X57482; CAA40720.1; -.
DR   EMBL; M88302; AAA39963.1; -.
DR   PIR; S11999; S11999.
DR   PIR; S13083; S13083.
DR   PIR; S30205; S30205.
DR   PIR; S31226; S31226.
DR   HSSP; P14859; 1OCT.
DR   TRANSFAC; T00656; -.
DR   MGD; MGI:101896; Pou3f1.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000327; POU_domain.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00157; pou; 1.
DR   PRINTS; PR00028; POUDOMAIN.
DR   ProDom; PD000010; Homeobox; 1.
DR   ProDom; PD000583; POU_domain; 1.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00352; POU; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS00035; POU_1; 1.
DR   PROSITE; PS00465; POU_2; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation; Homeobox.
FT   DOMAIN      249    319       POU.
FT   DNA_BIND    337    396       HOMEOBOX.
FT   CONFLICT     35     35       MISSING (IN REF. 2).
SQ   SEQUENCE   449 AA;  45323 MW;  82DD1335A0C4F43C CRC64;

  Query Match             10.9%;  Score 83;  DB 1;  Length 449;
  Best Local Similarity   26.4%;  Pred. No. 4.9;
  Matches   46;  Conservative   10;  Mismatches   76;  Indels   42;  Gaps    8;

Qy          8 PGAGNVSAAGLDPTEHPLLGATLELATD---GGALLAGRLSLRSHPWL---ADHAV---- 57
              || |   | |  |  ||   |    | :    ||        |  | ||   | | |
Db         12 PGGG---AGGTGPLMHPDAAAAAAAAAERLHAGAAYREVQKLMHHEWLGAGAGHPVGLAH 68

Qy         58 --------GGTVLLSGATFLELALHAGTYVGCDRVDE----LTLHAPLV---------VP 96
                     ||    :|   ||      |   |  | |:       || ||
Db         69 PQWLPTGGGGGGDWAGGPHLEHGKAGGG--GTGRADDGGGGGGFHARLVHQGAAHAGAAW 126

Qy         97 VDGGVSVQVGVA------AADGEGRRLVSVYARGGSACGGGGASGGVWTCHASG 144
                || :  :| |      |  |   : : :||:     ||||   |:      |
Db        127 AQGGTAHHLGPAMSPSPGAGGGHQPQPLGLYAQAAYPGGGGGGLAGMLAAGGGG 180


RESULT 10
FABH_RHOCA
ID   FABH_RHOCA     STANDARD;      PRT;   324 AA.
AC   P30790;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   3-oxoacyl-[acyl-carrier-protein] synthase III (EC 2.3.1.41) (Beta-
DE   ketoacyl-ACP synthase III) (KAS III).
GN   FABH.
```

```
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92073365; PubMed=1961742;
RA   Toussaint B., Bosc C., Richaud P., Colbeau A., Vignais P.M.;
RT   "A mutation in a Rhodobacter capsulatus gene encoding an integration
RT   host factor-like protein impairs in vivo hydrogenase expression.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10749-10753(1991).
CC   -!- FUNCTION: CATALYZES THE CONDENSATION REACTION OF FATTY ACID
CC       SYNTHESIS BY THE ADDITION TO AN ACYL ACCEPTOR OF TWO CARBONS
CC       FROM MALONYL-ACP. KAS III CATALYZES THE FIRST CONDENSATION
CC       REACTION WHICH INITIATES FATTY ACID SYNTHESIS AND MAY THEREFORE
CC       PLAY A ROLE IN GOVERNING THE TOTAL RATE OF FATTY ACID PRODUCTION.
CC       POSSESSES BOTH ACETOACETYL-ACP SYNTHASE AND ACETYL TRANSACYLASE
CC       ACTIVITIES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: BELONGS TO THE FABH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84030; AAA26127.1; -.
DR   PIR; C41608; C41608.
DR   HSSP; P24249; 1HNK.
DR   InterPro; IPR004655; FabH.
DR   ProDom; PD000453; N-C_synthase; 1.
DR   TIGRFAMs; TIGR00747; fabH; 1.
KW   Fatty acid biosynthesis; Transferase; Acyltransferase;
KW   Multifunctional enzyme.
FT   ACT_SITE    114    114       BY SIMILARITY.
FT   ACT_SITE    251    251       BY SIMILARITY.
FT   ACT_SITE    281    281       BY SIMILARITY.
SQ   SEQUENCE   324 AA;  33732 MW;  7303B98E76E5A20E CRC64;

  Query Match            10.7%;  Score 81.5;  DB 1;  Length 324;
  Best Local Similarity  26.5%;  Pred. No. 4.7;
  Matches   36;  Conservative   15;  Mismatches   48;  Indels   37;  Gaps    6;

Qy          7 QPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGR------LSLRSHPWLADHAVGGT 60
              | | || :|   |  :    |    ||   | :|:|:      :   :   : | |  ||
Db         97 QAGLGNTNAFAFD-VQAVCAGFVFALANANGMILSGQAKRVLVIGAETFSRIMDWADRGT 155

Qy         61 VLL----SGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD--GEG 114
                :|    :||  || |  |||    ||                       |: |:|   :|
Db        156 CVLFGDGAGAVILEAAEGAGT--AADR---------------------GILASDLHSDG 191

Qy        115 RRLVSVYARGGSACGG 130
              |    :|  || :  |
Db        192 RYRELLYVDGGVSTNG 207


RESULT 11
YI18_MYCTU
ID   YI18_MYCTU     STANDARD;      PRT;   498 AA.
AC   Q50615;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1818c.
GN   RV1818C OR MT1866 OR MTCY1A11.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z78020; CAB01461.1; -.
DR   EMBL; AE007045; AAK46139.1; ALT_INIT.
DR   TIGR; MT1866; -.
DR   TubercuList; Rv1818c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM     93    113       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    182    202       POTENTIAL.
FT   TRANSMEM    209    229       POTENTIAL.
FT   TRANSMEM    244    264       POTENTIAL.
FT   TRANSMEM    281    301       POTENTIAL.
FT   TRANSMEM    311    331       POTENTIAL.
FT   TRANSMEM    338    358       POTENTIAL.
FT   TRANSMEM    373    393       POTENTIAL.
FT   TRANSMEM    403    423       POTENTIAL.
FT   CONFLICT    183    212       MISSING (IN REF. 2).
FT   CONFLICT    414    414       G -> GAGG (IN REF. 2).
SQ   SEQUENCE   498 AA;  40755 MW;  4F6F78F2482586BA CRC64;

  Query Match            10.5%;  Score 80.5;  DB 1;  Length 498;
  Best Local Similarity  26.6%;  Pred. No. 8.9;
  Matches   38;  Conservative   13;  Mismatches   67;  Indels   25;  Gaps    5;

Qy         11 GNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLE 70
              |  :||   ||   |  |  |::    || :|          |  |:     |    |:
Db         26 GTANAAAAVPTTTVLAAAADEVSAAMAALFSGHAQ-------AYQALSAQAALFHEQFVR 78
```

```
Qy          71 -LALHAGTYVGCDRVDELTLHAPLVV----------PVDG-GVSVQVGVAAADGEGRRL 117
               |   ||:|   :       |   | |          |: | | :   |  |  |:|  |
Db          79 ALTAGAGSYAAAEAASAAPLEGVLDVINAPALALLGRPLIGNGANGAPGTGANGGDGGIL 138

Qy         118 VSVYARGGSAC-----GGGGASG 135
               :     |||       | |||:|
Db         139 IGNGGAGGSGAAGMPGGNGGAAG 161


RESULT 12
WA22_MYCTU
ID   WA22_MYCTU     STANDARD;      PRT;   914 AA.
AC   O06794;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   WAG22 antigen precursor.
GN   WAG22 OR RV1759C OR MT1807 OR MTCY28.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 85.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95890; CAB09322.1; -.
DR   EMBL; AE007040; -; NOT_ANNOTATED_CDS.
DR   TIGR; MT1807; -.
DR   TubercuList; Rv1759c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Antigen; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    914       WAG22 ANTIGEN.
SQ   SEQUENCE   914 AA;  74354 MW;  F6953C3DBE8E6AC8 CRC64;

 Query Match             10.5%;  Score 80.5;  DB 1;  Length 914;
  Best Local Similarity   27.9%;  Pred. No. 17;
  Matches   38;  Conservative    9;  Mismatches   60;  Indels   29;  Gaps    4;

Qy           9 GAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATF 68
               | |   |||  : :     ||   |   |||  ||  :             ||     |: |
Db         557 GNGGAGAAGTNGSAGGAGGAGGILFGTGGAGGAGGVGTA--------GAGGAGGAGGSAF 608

Qy          69 LELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSAC 128
               |   : :|   |                 | |  :   ||  | |    |:     ||  |
Db         609 L---IGSGGTGG---------------VGGAATTTGGVGGAGGNAGLLIGAAGLGG--C 647

Qy         129 GGGGASGGVWTCHASG 144
               |||  : || |  | |
Db         648 GGGAFTAGVTTGGAGG 663


RESULT 13
MB48_MYCTU
ID   MB48_MYCTU     STANDARD;      PRT;   460 AA.
AC   Q933K8;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Antigen MTB48.
GN   MTB48 OR RV3881C OR MT3996 OR MTV027.16C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Erdmann;
RX   MEDLINE=21320746; PubMed=11427558;
RA   Lodes M.J., Dillon D.C., Mohamath R., Day C.H., Benson D.R.,
RA   Reynolds L.D., McNeill P., Sampaio D.P., Skeiky Y.A.W., Badaro R.,
RA   Persing D.H., Reed S.G., Houghton R.L.;
RT   "Serological expression cloning and immunological evaluation of MTB48,
RT   a novel Mycobacterium tuberculosis antigen.";
RL   J. Clin. Microbiol. 39:2485-2493(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. A processed form is shed into
CC       the medium during culture.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AY029285; AAK31576.1; -.
DR   EMBL; AL022120; CAA17973.1; -.
DR   EMBL; AE007191; AAK48364.1; -.
DR   TIGR; MT3996; -.
DR   TuberculList; Rv3881c; -.
KW   Antigen; Complete proteome.
FT   DOMAIN      306    428       GLY-RICH.
SQ   SEQUENCE   460 AA;  47593 MW;  7528743226AD7A71 CRC64;

  Query Match             10.3%;  Score 79;  DB 1;  Length 460;
  Best Local Similarity   29.2%;  Pred. No. 11;
  Matches   40;  Conservative   11;  Mismatches   54;  Indels   32;  Gaps    6;

Qy          8 PGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGAT 67
              || |  :|  : ||  |  |  |    |       |||  :           |: | | :  |
Db        311 PGTGMPAAPMVPPTGSP--GGGLPADTAAQLTSAGREAA---------ALSGDVAVKAA- 358

Qy         68 FLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ---VGVAAADGEGRRLVSVYARG 124
                :|  |  |          |||  : |  ||:    |  |  |:||      | |
Db        359 ----SLGGGGGGGVP-------SAPLGSAIGGAESVRPAGAGDIAGLGQGR------AGG 401

Qy        125 GSACGGGGASGGVWTCH 141
              |:| ||||    :   |
Db        402 GAALGGGGMGMPMGAAH 418


RESULT 14
DEOC_MYCLE
ID   DEOC_MYCLE     STANDARD;      PRT;   226 AA.
AC   Q9CB45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
DE   (Deoxyriboaldolase).
GN   DEOC OR ML2451.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC       glyceraldehyde 3-phosphate + acetaldehyde.
CC   -!- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.
CC       DEOC SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL583925; CAC31968.1; -.
DR   Leproma; ML2451; -.
DR   InterPro; IPR002915; DeoC.
DR   InterPro; IPR003009; FMN_enzyme.
DR   Pfam; PF01791; DeoC; 1.
KW   Lyase; Schiff base; Complete proteome.
FT   BINDING     161    161       SCHIFF-BASE (BY SIMILARITY).
SQ   SEQUENCE   226 AA;  22527 MW;  2B79A2D09B84F8A4 CRC64;

  Query Match             10.1%;  Score 77.5;  DB 1;  Length 226;
  Best Local Similarity   27.0%;  Pred. No. 7.2;
  Matches   44;  Conservative   15;  Mismatches   59;  Indels   45;  Gaps    8;

Qy          5 STQPGAGNVSAA---GLDPTEHPLL-----------GAT-LELATDGGALLAGRL-SLR 47
              : : |||  :   |    |    ||  :           |||  :::    || ||| | ::|
Db         53 AVEAGAGGLRVATVVGFPSGEHVSIVKAQEAVQAVASGATEIDMVIAVGAALAGDLDAVR 112

Qy         48 SHPWLADHAVGGTVL---LSGATFLELA--------LHAGTYVGCDRVDELTLHAPLVVP 96
              |    :  |: | ||   :  | |  |         |     | | |   |    |
Db        113 SDIEMVRSAIAGVVLKVIVESAVLLGQANERTLAEVCHVAEAAGADFVKTSTGFHPA--- 169

Qy         97 VDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWT 139
                || ||:      :  | ||           |  ||||: |
Db        170 --GGASVRAVTLMTEAVGGRL-----------GVKASGGIRT 198


RESULT 15
YM05_MYCTU
ID   YM05_MYCTU     STANDARD;      PRT;   358 AA.
AC   Q10394;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein Rv2205c.
GN   RV2205C OR MT2261 OR MTCY190.16C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z70283; CAA94248.1; ALT_INIT.
DR   EMBL; AE007072; AAK46547.1; ALT_INIT.
DR   TIGR; MT2261; -.
DR   TubercuList; Rv2205c; -.
DR   InterPro; IPR004381; Cons_hypoth45.
DR   InterPro; IPR003747; Glycerate_kinase.
DR   Pfam; PF02595; DUF168; 1.
DR   TIGRFAMs; TIGR00045; Cons_hypoth45; 2.
KW   Hypothetical protein; Transferase; Kinase; Complete proteome.
SQ   SEQUENCE   358 AA;  35592 MW;  AC7EDC5BF8E41544 CRC64;

  Query Match             10.1%;  Score 77.5;  DB 1;  Length 358;
  Best Local Similarity   25.5%;  Pred. No. 11;
  Matches   39;  Conservative   19;  Mismatches   58;  Indels   37;  Gaps    7;

Qy          5 STQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLS 64
              | :||||   |||        |    |   ||   :|   :  |   ||       ::::
Db        219 SAEPGAG---AAG---------GIGAGLLAVGGRYQSGAAIIAEHTHFADDLADAELIVT 266

Qy         65 G-ATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSV------------------- 103
              |   | | :|| |  ||         |  |::| : | ||:
Db        267 GEGRFDEQSLH-GKVVGAIAAAARPLAIPVIV-LAGQVSLDKSALRSAGIMAALSIAEYA 324

Qy        104 -QVGVAAADGEGRRLVSVYARGGSACGGGGASG 135
               | :| ||    :|: : ::  :  |  | ||
Db        325 GSVRLALADA-ANQLMGLASQVAARLGNSGPSG 356


Search completed: June 17, 2003, 13:02:39
Job time : 3.59218 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 3.26017 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_2983_3128
Perfect score: 764
Sequence: 1 YWLESTQPGAGNVSAAGLDP.....ACGGGGASGGVWTCHASGVL 146

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	385.5	50.5	3739	2	T17410	polyketide synthas
2	361	47.3	6420	2	T30283	polyketide synthas
3	346.5	45.4	2103	2	G86925	probable polyketid
4	332	43.5	5069	2	T17464	rifamycin polyketi
5	329	43.1	10223	2	T30225	polyketide synthas
6	324.5	42.5	4151	2	G70944	probable polyketid
7	320	41.9	4613	2	T17409	polyketide synthas
8	317	41.5	7576	2	T17428	PK506 polyketide s
9	315.5	41.3	1937	2	T03224	probable polyketid
10	314.5	41.2	8563	2	T30226	polyketide synthas
11	313.5	41.0	2124	2	T28658	polyketide synthas
12	312	40.8	2126	2	H70621	probable polyketid
13	308.5	40.4	3413	2	T17467	rifamycin polyketi
14	304	39.8	1616	2	G70668	polyketide synthas
15	300	39.3	1763	2	T17465	rifamycin polyketi
16	299	39.1	1602	2	H70984	probable polyketid
17	279.5	36.6	6260	2	T30228	polyketide synthas
18	271.5	35.5	1762	2	T03222	probable polyketid
19	245.5	32.1	2100	2	T03223	probable polyketid
20	240.5	31.5	2723	2	T03221	probable polyketid
21	234	30.6	3573	2	S23070	erythronolide synt
22	201.5	26.4	1728	2	T17466	rifamycin polyketi
23	147.5	19.3	4735	2	T17463	rifamycin polyketi
24	126.5	16.6	2188	2	A70984	probable polyketid
25	121	15.8	1827	2	B70984	probable polyketid
26	103	13.5	2201	2	S73014	polyketide synthas
27	101.5	13.3	2108	2	H70819	probable polyketid
28	99.5	13.0	2118	2	S72705	mycocerosate synth
29	97.5	12.8	2111	2	A70668	mycocerosate synth
30	97.5	12.8	2126	2	E70522	probable polyketid
31	96.5	12.6	2110	2	B44110	mycocerosate synth
32	95	12.4	741	2	G70917	hypothetical glyci
33	95	12.4	2116	2	C86926	probable mycoceros
34	92	12.0	1582	2	E70876	probable polyketid
35	91.5	12.0	2518	2	AI2140	polyketide synthas
36	91	11.9	1815	2	S73021	polyketide synthas
37	91	11.9	1822	2	F87203	polyketide synthas
38	89	11.6	433	2	F75566	probable benzoate
39	88.5	11.6	775	2	T07172	subtilisin-like pr
40	88	11.5	2174	2	E95965	hypothetical glyci
41	88	11.5	2554	2	AB3528	extracellular seri
42	87.5	11.5	1028	2	T03516	probable outer mem
43	86.5	11.3	509	2	T34871	probable membrane
44	86	11.3	1027	2	F87370	alpha-L-rhamnosida
45	86	11.3	4936	2	AH2515	hypothetical prote

ALIGNMENTS

RESULT 1
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

Query Match 50.5%; Score 385.5; DB 2; Length 3739;
Best Local Similarity 54.8%; Pred. No. 5.3e-25;
Matches 80; Conservative 14; Mismatches 47; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

RESULT 2
T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30283
R;Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A;Title: Structural organization of a multifunctional polyketide synthase involved in t
A;Reference number: Z20806; MUID:97217427; PMID:9063448
A;Accession: T30283
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA
A;Residues: 1-6420 <MOT>
A;Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C;Genetics:
A;Note: fkbA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3391-3462/Domain: acyl carrier protein homology <ACP1>
F;3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5307-5378/Domain: acyl carrier protein homology <ACP2>
F;5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6293-6364/Domain: acyl carrier protein homology <ACP3>

Query Match 47.3%; Score 361; DB 2; Length 6420;
Best Local Similarity 50.0%; Pred. No. 1.3e-22;
Matches 73; Conservative 18; Mismatches 45; Indels 10; Gaps 2;

Qy 1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
Db 4309 YWIEAT--GAADLTALGLTDTAHPMLAATAALPTNDGTVLTGRISLRTHPWLADHTIQDT 4366

Qy 61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
Db 4367 VLLPGTAFLELVMRAGDEVGCDTIDELVIETPLTLPVTGAVDLTVTVEQADATGHRPVSV 4426

Qy 121 YARGGSACGGGGASGGVWTCHASGVL 146
Db 4427 HAR--------PAGTDTWTRHATGTL 4444

RESULT 3
G86925
probable polyketide synthase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 24-May-2001
C;Accession: G86925
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: G86925
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2103 <STO>
A;Cross-references: GB:AL450380; NID:g13092513; PIDN:CAC29643.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0135
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein

Query Match 45.4%; Score 346.5; DB 2; Length 2103;
Best Local Similarity 50.7%; Pred. No. 6.9e-22;
Matches 74; Conservative 14; Mismatches 51; Indels 7; Gaps 1;

Qy 1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
Db 909 FWETPGTDGTASVSGLGLGSTEHALLGAVVERPDSGGVVLTGRLSLANQPWLADHVIGGV 968

Qy 61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
Db 969 VLFPGAGFVELVIRAGDEVGCAGVEELVLAAPLVLHPGSGVQVQVVVAAADESSRRAVSV 1028

Qy 121 YARGGSACGGGGASGGVWTCHASGVL 146
Db 1029 YSRGDQSHGG-------WLLNAEGLL 1047

RESULT 4
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 43.5%; Score 332; DB 2; Length 5069;
Best Local Similarity 47.3%; Pred. No. 3.2e-20;
Matches 69; Conservative 20; Mismatches 49; Indels 8; Gaps 3;

Qy 1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
Db 4195 YWLRPAE-SATDAASLGQAAADHPLLGAVVELPQSDGLVFTSRLSVRTHPWLADHAVGGV 4253

Qy 61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
Db 4254 VILPGSGLAELAVRAGDEAGCTALDELIIEAPLVVPAQGAVRVQVALSGPDETGSRTVDL 4313

Qy 121 YARGGSACGGGGASGGVWTCHASGVL 146
Db 4314 YSQ-----RDGGA--GTWTRHATGVL 4332

RESULT 5
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>

```
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match             43.1%;  Score 329;  DB 2;  Length 10223;
  Best Local Similarity   47.3%;  Pred. No. 1.2e-19;
  Matches   69;  Conservative   19;  Mismatches   48;  Indels   10;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || |   :  : :|| ||||   |||||| : |    | || ||:|| :| ||||||| |:
Db       7685 YWAEVGR--SADVSGAGLDAVGHPLLGAVVALPGSDGVLLTGRVSLATHAWLADHAVRGS 7742

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|| .: |   |||| |||| : |||::|  ||| | | |  ||  | |:|::
Db       7743 VLLPGTGFVELVVRAADEVGCDVVDELIVEAPLLLPQTGGVQVSVSVGEADESGHRVVTI 7802

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ::|  :|          | | | |:
Db       7803 FSRADNA--------DTWVRHVSAVI 7820


RESULT 6
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

  Query Match             42.5%;  Score 324.5;  DB 2;  Length 4151;
  Best Local Similarity   47.3%;  Pred. No. 1.2e-19;
  Matches   69;  Conservative   18;  Mismatches   50;  Indels    9;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              :|| :   || :||  ||  :|||||||| ::|   || :| ||||    ||||||||
Db        905 FWLSAEGSGA-DVSGLGLGASEHPLLGAVVDLPASGGVVLTGRLSPNVQPWLADHAVSDV 963

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||  |  |:||| || |||  :||||| |||::|  | |:||| | |       | ||:
Db        964 VLPPGTGFVELAIRAGDEVGCSVLDELTLAAPLLLPATGSVAVQVVVDAGRDSNSRGVSI 1023

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ::|  :  |         |  || |:|
Db       1024 FSRADAQAG--------WLLHAEGIL 1041


RESULT 7
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

  Query Match             41.9%;  Score 320;  DB 2;  Length 4613;
  Best Local Similarity   47.1%;  Pred. No. 3.2e-19;
  Matches   74;  Conservative   19;  Mismatches   46;  Indels   18;  Gaps    4;

Qy          1 YWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              |||::  | |    : :  ||   :|||||| : |    | || |||||||:|||||||||
Db       3556 YWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRLSLRTHPWLADHAV 3613

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE---- 113
              |:||| ||  :||| ||     |   | |||  |||:|  ||| ::| | |  ||
Db       3614 LGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELRVTVGAPAGEPGGE 3673

Qy        114 ----GRRLVSVYARGGSACGGGGASGGVWTCHASGVL 146
                  | | ||:::|   |      :|  |:|||:|:|
Db       3674 SAGDGARPVSLHSRLADA-----PAGTAWSCHATGLL 3705


RESULT 8
T17428
FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: T17428
R;Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressa
A;Reference number: Z18779; MUID:98451508; PMID:9780228
A;Accession: T17428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7576 <MOT>
A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A;Experimental source: strain MA6548
C;Genetics:
A;Gene: fkbB
C;Function:
A;Description: involved in synthesis of the backbone of the immunosuppressant FK506 pol
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-500/Domain: acetate-CoA ligase homology <ACL>
F;1095-1166/Domain: acyl carrier protein homology <ACP1>
F;1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;2680-2751/Domain: acyl carrier protein homology <ACP2>
F;2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;4320-4391/Domain: acyl carrier protein homology <ACP3>
F;4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5903-5974/Domain: acyl carrier protein homology <ACP4>
F;6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;7450-7521/Domain: acyl carrier protein homology <ACP5>
```

```
  Query Match             41.5%;  Score 317;  DB 2;  Length 7576;
  Best Local Similarity   47.6%;  Pred. No. 9.8e-19;
  Matches   70;  Conservative   14;  Mismatches   49;  Indels   14;  Gaps    4;

Qy          1 YWLESTQPGAG-NVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 59
              ||:|   |  | : |  |||  :|||| |:: |      || ||||| :|||| | | |
Db       6851 YWIE---PATGTDASGLGLDAVDHPLLAASVALPDSDEVLLTGRLSLATHPWLAGHTVDG 6907

Qy         60 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 119
               ||| |  |:|||       || :||| :  || :|  | | |:| |||||  ||| |
Db       6908 EVLLPGPVFVELAGRGADEARCDLLDELVIETPLALPATGAVQVRVTVAAADDAGRRAVR 6967

Qy        120 VYARGGSACGGGGASGG-VWTCHASGV 145
              :||          | ||  || ||||:
Db       6968 IHAR---------AHGGRTWTRHASGI 6985


RESULT 9
T03224
probable polyketide synthase module 4 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03224
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03224
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1937 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38064.1; PID:g2624951
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;549-827/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

  Query Match             41.3%;  Score 315.5;  DB 2;  Length 1937;
  Best Local Similarity   45.9%;  Pred. No. 3.1e-19;
  Matches   67;  Conservative   18;  Mismatches   50;  Indels   11;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||      |: :  || || :|||||   :|  |  :|:||||: | |||||||| ||
Db        884 YWLERGSGSGGDAAGLGLLPTRYPLLGAVTTIAGSGAVVLSGRLSVASMPWLADHAVAGT 943

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              |:| |   ::: : ||  |||| :||| :  |||  :|   |||| |   : :||| |||
Db        944 VVLPGTALVDMVIRAGDEVGCDVLDELVVQTPLV--LDRAASVQVSVGPTEADGRRPVSV 1001

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :||           | |  || | |
Db       1002 HAR---------RDDGDWVEHAVGSL 1018


RESULT 10
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             41.2%;  Score 314.5;  DB 2;  Length 8563;
  Best Local Similarity   44.5%;  Pred. No. 1.8e-18;
  Matches   65;  Conservative   18;  Mismatches   46;  Indels   17;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||::|    | :          ||||||  :||    | || ||:|| :| |||||||  |
Db       7692 YWVKSVDRAAAD---------GHPLLGAVVELPESDGVLLTGRVSLATHAWLADHAVWGR 7742

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|| :||   ||||  |||| :  ||::|  ||| : | |  ||  | |:|:|
Db       7743 VLLPGTAFVELVVHAAGEVGCDVVDELVIETPLLLPQTGGVQLSVSVGEADESGHRVVTV 7802

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ::|  :|          || | |  :
Db       7803 FSRADNA--------DTWTRHVSATV 7820


RESULT 11
T28658
polyketide synthase - Sorangium cellulosum (fragment)
C;Species: Sorangium cellulosum
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C;Accession: T28658
R;Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J.J.; Ligon, J.M.
J. Bacteriol. 177, 3673-3679, 1995
A;Title: A Sorangium cellulosum (myxobacterium) gene cluster for the biosynthesis of th
inomycetes.
A;Reference number: Z20499; MUID:95325306; PMID:7601830
A;Accession: T28658
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2124 <SCH>
A;Cross-references: EMBL:U24241; NID:g868150; PID:g868151; PIDN:AAA79984.1
A;Experimental source: strain So ce26
C;Function:
A;Description: involved in soraphen A biosynthesis
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;335-406/Domain: acyl carrier protein homology <ACP>
F;370/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             41.0%;  Score 313.5;  DB 2;  Length 2124;
  Best Local Similarity   45.2%;  Pred. No. 5.1e-19;
  Matches   66;  Conservative   20;  Mismatches   55;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              :||:::   | :|::|||   :|||||| : ||   | :  |||||  |||| || | |
Db       1314 FWLDASTAHAADVASAGLTSADHPLLGAAVALADRDGFVFTGRLSLAEHPWLEDHVVFGI 1373

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
                 :   ||||||   || | |:::||  || :|  | | :|: |  ||| ||| :||
Db       1374 PCPARRRLLELALHVAHLVGLDTVEDVTLDPPLALPSQGAVLLQISVGPADGAGRRALSV 1433

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ::|   |        | || |||| |
Db       1434 HSRRHDA-----LQDGPWTRHASGSL 1454
```

```
RESULT 12
H70621
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70621
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2126 <COL>
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06632.1; PID:g1839013
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks7
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;563-845/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1694-1875/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1976-2051/Domain: acyl carrier protein homology <ACP2>

  Query Match             40.8%;  Score 312;  DB 2;  Length 2126;
  Best Local Similarity   47.3%;  Pred. No. 6.8e-19;
  Matches   69;  Conservative   15;  Mismatches   56;  Indels    6;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              :||:      |::   ||   || || | ::    || :| ||:|: : |||||||||
Db        912 FWLDGLGAVGGDLGGVGLVGAEHGLLAAVVQRPDSGGVVLTGRISVVAAPWLADHAVGPV 971

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||  |  |:|||| ||  |||  : |||| ||||:| | || ||| |   :  | | | |
Db        972 VLFPGTGFVELALRAGDEVGCSVLQELTLQAPLVLPAD-GVRVQVVVGGVEQSGTRNVWV 1030

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |:  | |    |     || || |||
Db       1031 YSAAGQADSSPG-----WTLHAQGVL 1051


RESULT 13
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

  Query Match             40.4%;  Score 308.5;  DB 2;  Length 3413;
  Best Local Similarity   47.3%;  Pred. No. 2.3e-18;
  Matches   70;  Conservative   15;  Mismatches   52;  Indels   11;  Gaps    4;

Qy          1 YWLESTQPGAGNVSAAGLD--PTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVG 58
              |||   || |    || |     :|||||| : |    | :   ||||:||||||||| :|

Db        897 YWL---QPAAQATDAASLGQVAADHPLLGAVVRLPQSDGLVFTSRLSLKSHPWLADHVIG 953

Qy         59 GTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLV 118
              | ||::|   :|||: ||   ||  ::|| : ||||||  ||| :|| | |    | | |
Db        954 GVVLVAGTGLVELAVRAGDEAGCPVLEELVIEAPLVVPDHGGVRIQVVVGAPGETGSRAV 1013

Qy        119 SVYARGGSACGGGGASGGVWTCHASGVL 146
              ||:    |    ||   ||  ||:| |
Db       1014 EVYSLREDA----GAE--VWARHATGFL 1035


RESULT 14
G70668
polyketide synthase pks1 - Mycobacterium tuberculosis (strain H37RV)
N;Alternate names: polyketide synthase pks002c
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70668; S73075
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70668
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1616 <COL>
A;Cross-references: GB:Z83858; GB:AL123456; NID:g3261675; PIDN:CAB06103.1; PID:g1781167
A;Experimental source: strain H37Rv
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A;Description: Mycobacterium tuberculosis cosmid tbc2.
A;Reference number: S73053
A;Accession: S73075
A;Molecule type: DNA
A;Residues: 'MWGARWPVARCLSTEL',48-917,'RSEARSRRGCGMSGRAAPPAPPIGPPIGIP','PMPPIIPVEPAGPVF
A;Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50928.1; PID:g560507
C;Genetics:
A;Gene: pks1
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase pks1; acyl carri
ology
C;Keywords: carrier protein
F;84-362/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1228-1410/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1514-1585/Domain: acyl carrier protein homology <ACP1>

  Query Match             39.8%;  Score 304;  DB 2;  Length 1616;
  Best Local Similarity   45.9%;  Pred. No. 2.5e-18;
  Matches   67;  Conservative   14;  Mismatches   57;  Indels    8;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              :|      |  : :  ||  ||| |||| :|       :| |||||   |||||| | |
Db        423 FWETPGADGPADAAGLGLGATEHALLGAVVERPDSDEVVLTGRLSLADQPWLADHVVNGV 482

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||  || |:|| : ||  |||  ::|| | ||||:    || ||| | |||  | | |||
Db        483 VLFPGAGFVELVIRAGDEVGCALIEELVLAAPLVMHPGVGVQVQVVVGAADESGHRAVSV 542

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |:||  : |        |  :| |:|
Db        543 YSRGDQSQG--------WLLNAEGML 560


RESULT 15
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
```

```
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>

  Query Match            39.3%;  Score 300;  DB 2;  Length 1763;
  Best Local Similarity  43.5%;  Pred. No. 6.1e-18;
  Matches   64;  Conservative   19;  Mismatches   50;  Indels   14;  Gaps    3;

Qy          1 YWLESTQPGA-GNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 59
              :||   :| |  :  : |    ||||||| : |    | :   |||||:|||||||| :||
Db        891 FWL---RPAAQADAVSLGQAAAEHPLLGAVVRLPQSDGLVFTSRLSLRTHPWLADHTIGG 947

Qy         60 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 119
               ||  |   :|||: ||   ||  :|||   ||||||  |||:||| |:  |  | | |
Db        948 VVLFPGTGLVELAVRAGDEAGCPVLDELVTEAPLVVPGQGGVNVQVTVSGPDQNGLRTVD 1007

Qy        120 VYARGGSACGGGGASGGVWTCHASGVL 146
              ::::             ||| ||:| :
Db       1008 IHSQ----------RDDVWTRHATGTV 1024


Search completed: June 17, 2003, 13:12:24
Job time : 4.26017 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 4.54907 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_2983_3128
Perfect score: 764
Sequence: 1 YWLESTQPGAGNVSAAGLDP.....ACGGGGASGGVWTCHASGVL 146

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	385.5	50.5	3739	9	US-09-860-846-33	Sequence 33, Appl
2	385.5	50.5	3739	9	US-09-988-384B-33	Sequence 33, Appl
3	385.5	50.5	3739	9	US-09-836-821-33	Sequence 33, Appl
4	385.5	50.5	3739	9	US-09-793-708-2	Sequence 2, Appli
5	385.5	50.5	3739	10	US-09-861-289-33	Sequence 33, Appl
6	385.5	50.5	11877	9	US-09-860-846-6	Sequence 6, Appli
7	385.5	50.5	11877	9	US-09-836-821-6	Sequence 6, Appli
8	385.5	50.5	11877	10	US-09-861-289-6	Sequence 6, Appli
9	385.5	50.5	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	324	42.4	5215	9	US-09-860-846-2	Sequence 2, Appli
11	324	42.4	5215	9	US-09-988-384B-2	Sequence 2, Appli
12	324	42.4	5215	9	US-09-836-821-2	Sequence 2, Appli
13	324	42.4	5215	10	US-09-861-289-2	Sequence 2, Appli
14	320	41.9	4551	9	US-09-793-708-1	Sequence 1, Appli
15	320	41.9	4613	9	US-09-860-846-31	Sequence 31, Appl
16	320	41.9	4613	9	US-09-988-384B-31	Sequence 31, Appl
17	320	41.9	4613	9	US-09-836-821-31	Sequence 31, Appl
18	320	41.9	4613	10	US-09-861-289-31	Sequence 31, Appl
19	319	41.8	3816	9	US-09-808-880-3	Sequence 3, Appli
20	304	39.8	1616	9	US-09-712-363-262	Sequence 262, App
21	155	20.3	1832	9	US-10-014-717-4	Sequence 4, Appli
22	155	20.3	7257	9	US-10-014-717-5	Sequence 5, Appli
23	154	20.2	3798	9	US-10-014-717-6	Sequence 6, Appli
24	147	19.2	2439	9	US-10-014-717-7	Sequence 7, Appli
25	121	15.8	1827	9	US-09-712-363-261	Sequence 261, App
26	118.5	15.5	2563	9	US-09-836-705-46	Sequence 46, Appl
27	96	12.6	1674	9	US-10-123-155-453	Sequence 453, App
28	85	11.1	256	9	US-09-880-748-1607	Sequence 1607, Ap
29	84.5	11.1	423	10	US-09-073-009-142	Sequence 142, App
30	84.5	11.1	423	10	US-09-793-306-142	Sequence 142, App
31	84.5	11.1	462	9	US-10-270-333-33	Sequence 33, Appl
32	84.5	11.1	710	10	US-09-287-849-16	Sequence 16, Appl
33	84.5	11.1	856	10	US-09-287-849-12	Sequence 12, Appl
34	83	10.9	1567	9	US-10-123-155-215	Sequence 215, App
35	82.5	10.8	1797	9	US-10-184-644-331	Sequence 331, App
36	82.5	10.8	1797	9	US-10-184-634-331	Sequence 331, App
37	81.5	10.7	1094	9	US-10-184-644-387	Sequence 387, App
38	81.5	10.7	1094	9	US-10-184-634-387	Sequence 387, App
39	81.5	10.7	1621	9	US-10-184-644-145	Sequence 145, App
40	81.5	10.7	1621	9	US-10-184-634-145	Sequence 145, App
41	80.5	10.5	2698	9	US-10-123-155-25	Sequence 25, Appl
42	80.5	10.5	4374	9	US-10-123-155-125	Sequence 125, App
43	79.5	10.4	251	9	US-09-880-748-1510	Sequence 1510, Ap
44	79.5	10.4	1264	9	US-10-123-155-1	Sequence 1, Appli
45	79.5	10.4	1857	9	US-10-123-155-365	Sequence 365, App

ALIGNMENTS

RESULT 1
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33

Query Match 50.5%; Score 385.5; DB 9; Length 3739;
Best Local Similarity 54.8%; Pred. No. 4.2e-26;
Matches 80; Conservative 14; Mismatches 47; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |         || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

RESULT 2
US-09-988-384B-33

```
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             50.5%;  Score 385.5;  DB 9;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 4.2e-26;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546


RESULT 3
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             50.5%;  Score 385.5;  DB 9;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 4.2e-26;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546


RESULT 4
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-2

  Query Match             50.5%;  Score 385.5;  DB 9;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 4.2e-26;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546


RESULT 5
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             50.5%;  Score 385.5;  DB 10;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 4.2e-26;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||    | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546


RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match             50.5%;  Score 385.5;  DB 9;  Length 11877;
  Best Local Similarity   54.8%;  Pred. No. 1.6e-25;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||    | || | ||||:||||||||| ||
Db       7355 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 7414

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       7415 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 7474

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       7475 YAHPEDAPG-----EAEWTRHATGVL 7495


RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-6

  Query Match             50.5%;  Score 385.5;  DB 9;  Length 11877;
  Best Local Similarity   54.8%;  Pred. No. 1.6e-25;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||    | || | ||||:||||||||| ||
Db       7355 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 7414

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       7415 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 7474

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       7475 YAHPEDAPG-----EAEWTRHATGVL 7495


RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             50.5%;  Score 385.5;  DB 10;  Length 11877;
  Best Local Similarity   54.8%;  Pred. No. 1.6e-25;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||    | || | ||||:||||||||| ||
Db       7355 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 7414

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       7415 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 7474

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       7475 YAHPEDAPG-----EAEWTRHATGVL 7495


RESULT 9
US-09-988-384B-6
```

```
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             50.5%;  Score 385.5;  DB 9;  Length 12199;
  Best Local Similarity   54.8%;  Pred. No. 1.6e-25;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||:::.:|||   ||||||| : ||   | || | ||||:|||||||| ||
Db       7677 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 7736

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       7737 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 7796

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |         || ||:|||
Db       7797 YAHPEDAPG-----EAEWTRHATGVL 7817


RESULT 10
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-2

  Query Match             42.4%;  Score 324;  DB 9;  Length 5215;
  Best Local Similarity   52.0%;  Pred. No. 2.2e-20;
  Matches   77;  Conservative    9;  Mismatches   56;  Indels    6;  Gaps    4;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  |      :  | | ||  ||||  |:|:|   | || |||||   |||||| | |
Db       1926 YWL--TPEARTDARALGFDPARHPLLTTTVEVAGGDGVLLTGRLSLTDQPWLADHMVNGA 1983

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE--GRRLV 118
              |||    |||||| || :||  ||:|||| ||||:|  | | :|||| : |||     |
Db       1984 VLLPATAFLELALAAGDHVGAVRVEELTLEAPLVLPERGAVRIQVGV-SGDGESPAGRTF 2042

Qy        119 SVYARGGSACGGGGASGGVWTCHASGVL 146
              ||:   |   | |   || | ||||
Db       2043 GVYSTPDSGDTGDDAP-REWTRHVSGVL 2069


RESULT 11
US-09-988-384B-2
; Sequence 2, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-2

  Query Match             42.4%;  Score 324;  DB 9;  Length 5215;
  Best Local Similarity   52.0%;  Pred. No. 2.2e-20;
  Matches   77;  Conservative    9;  Mismatches   56;  Indels    6;  Gaps    4;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  |      :  | | ||  ||||  |:|:|   | || |||||  |||||| | |
Db       1926 YWL--TPEARTDARALGFDPARHPLLTTTVEVAGGDGVLLTGRLSLTDQPWLADHMVNGA 1983

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE--GRRLV 118
              |||    |||||| || :||  ||:|||| ||||:|  | | :|||| : |||     |
Db       1984 VLLPATAFLELALAAGDHVGAVRVEELTLEAPLVLPERGAVRIQVGV-SGDGESPAGRTF 2042

Qy        119 SVYARGGSACGGGGASGGVWTCHASGVL 146
              ||:   |   | |   || | ||||
Db       2043 GVYSTPDSGDTGDDAP-REWTRHVSGVL 2069


RESULT 12
US-09-836-821-2
; Sequence 2, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-2

  Query Match             42.4%;  Score 324;  DB 9;  Length 5215;
  Best Local Similarity   52.0%;  Pred. No. 2.2e-20;
  Matches   77;  Conservative    9;  Mismatches   56;  Indels    6;  Gaps    4;
```

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  |      :  | | ||  ||||  |:|:|   | || |||||   |||||| | |
Db       1926 YWL--TPEARTDARALGFDPARHPLLTTTVEVAGGDGVLLTGRLSLTDQPWLADHMVNGA 1983

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE--GRRLV 118
              |||    |||||| || :||  ||:|||| ||||:|  | | :|||| : |||    |
Db       1984 VLLPATAFLELALAAAGDHVGAVRVEELTLEAPLVLPERGAVRIQVGV-SGDGESPAGRTF 2042

Qy        119 SVYARGGSACGGGGASGGVWTCHASGVL 146
              ||:   |   |  |    || | ||||
Db       2043 GVYSTPDSGDTGDDAP-REWTRHVSGVL 2069


RESULT 13
US-09-861-289-2
; Sequence 2, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-2

  Query Match             42.4%;  Score 324;  DB 10;  Length 5215;
  Best Local Similarity   52.0%;  Pred. No. 2.2e-20;
  Matches   77;  Conservative    9;  Mismatches   56;  Indels    6;  Gaps    4;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  |      :  | | ||  ||||  |:|:|   | || |||||   |||||| | |
Db       1926 YWL--TPEARTDARALGFDPARHPLLTTTVEVAGGDGVLLTGRLSLTDQPWLADHMVNGA 1983

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE--GRRLV 118
              |||    |||||| || :||  ||:|||| ||||:|  | | :|||| : |||    |
Db       1984 VLLPATAFLELALAAAGDHVGAVRVEELTLEAPLVLPERGAVRIQVGV-SGDGESPAGRTF 2042

Qy        119 SVYARGGSACGGGGASGGVWTCHASGVL 146
              ||:   |   |  |    || | ||||
Db       2043 GVYSTPDSGDTGDDAP-REWTRHVSGVL 2069


RESULT 14
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-1

  Query Match             41.9%;  Score 320;  DB 9;  Length 4551;
  Best Local Similarity   47.1%;  Pred. No. 4.4e-20;
  Matches   74;  Conservative   19;  Mismatches   46;  Indels   18;  Gaps    4;

Qy          1 YWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              |||::  | |    : :  ||   :|||||| : |    | || |||||||:||||||||
Db       3494 YWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRLSLRTHPWLADHAV 3551

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE---- 113
              |:||| ||  :||| ||    |   | ||||  |||:|  ||| ::| | |  ||
Db       3552 LGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELRVTVGAPAGEPGGE 3611

Qy        114 ----GRRLVSVYARGGSACGGGGASGGVWTCHASGVL 146
                  | | ||:::|   |      :|  |:|||:|:|
Db       3612 SAGDGARPVSLHSRLADA-----PAGTAWSCHATGLL 3643


RESULT 15
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-31

  Query Match             41.9%;  Score 320;  DB 9;  Length 4613;
  Best Local Similarity   47.1%;  Pred. No. 4.5e-20;
  Matches   74;  Conservative   19;  Mismatches   46;  Indels   18;  Gaps    4;

Qy          1 YWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              |||::  | |    : :  ||   :|||||| : |    | || |||||||:||||||||
Db       3556 YWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRLSLRTHPWLADHAV 3613

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE---- 113
              |:||| ||  :||| ||    |   | ||||  |||:|  ||| ::| | |  ||
Db       3614 LGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELRVTVGAPAGEPGGE 3673

Qy        114 ----GRRLVSVYARGGSACGGGGASGGVWTCHASGVL 146
                  | | ||:::|   |      :|  |:|||:|:|
Db       3674 SAGDGARPVSLHSRLADA-----PAGTAWSCHATGLL 3705


Search completed: June 17, 2003, 13:46:46
```

Job time : 5.54907 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 2.65363 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_2983_3128
Perfect score: 764
Sequence: 1 YWLESTQPGAGNVSAAGLDP.....ACGGGGASGGVWTCHASGVL 146

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	385.5	50.5	3739	3	US-09-320-878-2	Sequence 2, Appli
2	385.5	50.5	3739	4	US-09-105-537-33	Sequence 33, Appl
3	385.5	50.5	11877	4	US-09-105-537-6	Sequence 6, Appli
4	365.5	47.8	5588	4	US-09-036-987A-6	Sequence 6, Appli
5	365.5	47.8	5588	4	US-09-370-700-6	Sequence 6, Appli
6	342	44.8	2152	4	US-09-036-987A-3	Sequence 3, Appli
7	342	44.8	2152	4	US-09-370-700-3	Sequence 3, Appli
8	333.5	43.7	4472	2	US-08-804-227C-2	Sequence 2, Appli
9	324	42.4	5215	4	US-09-105-537-2	Sequence 2, Appli
10	320	41.9	4551	3	US-09-320-878-1	Sequence 1, Appli
11	320	41.9	4613	4	US-09-105-537-31	Sequence 31, Appl
12	319	41.8	3816	4	US-09-428-517-3	Sequence 3, Appli
13	316.5	41.4	4545	2	US-08-804-227C-14	Sequence 14, Appl
14	316.5	41.4	4550	2	US-08-804-227C-8	Sequence 8, Appli
15	316.5	41.4	4550	2	US-08-804-198-2	Sequence 2, Appli
16	311.5	40.8	4928	4	US-09-036-987A-5	Sequence 5, Appli
17	311.5	40.8	4928	4	US-09-370-700-5	Sequence 5, Appli
18	308.5	40.4	3729	2	US-08-804-227C-4	Sequence 4, Appli
19	291	38.1	5087	4	US-09-144-085-1	Sequence 1, Appli
20	277.5	36.3	6095	4	US-09-144-085-2	Sequence 2, Appli
21	276.5	36.2	3724	2	US-08-804-227C-10	Sequence 10, Appl
22	276.5	36.2	3724	2	US-08-804-198-4	Sequence 4, Appli
23	265	34.7	1864	2	US-08-804-227C-3	Sequence 3, Appli
24	246.5	32.3	3567	2	US-07-642-734C-4	Sequence 4, Appli
25	246.5	32.3	3567	3	US-08-439-009A-4	Sequence 4, Appli
26	213	27.9	1996	2	US-08-804-227C-9	Sequence 9, Appli
27	213	27.9	1996	2	US-08-804-198-3	Sequence 3, Appli
28	155	20.3	1832	3	US-09-335-409-4	Sequence 4, Appli
29	155	20.3	1832	4	US-09-568-102-4	Sequence 4, Appli
30	155	20.3	1832	4	US-09-567-969-4	Sequence 4, Appli
31	155	20.3	1832	4	US-09-568-480-4	Sequence 4, Appli
32	155	20.3	1832	4	US-09-568-486-4	Sequence 4, Appli
33	155	20.3	1832	4	US-09-568-472-4	Sequence 4, Appli
34	155	20.3	1832	4	US-09-567-899-4	Sequence 4, Appli
35	155	20.3	7257	3	US-09-335-409-5	Sequence 5, Appli
36	155	20.3	7257	4	US-09-568-102-5	Sequence 5, Appli
37	155	20.3	7257	4	US-09-567-969-5	Sequence 5, Appli
38	155	20.3	7257	4	US-09-568-480-5	Sequence 5, Appli
39	155	20.3	7257	4	US-09-568-486-5	Sequence 5, Appli
40	155	20.3	7257	4	US-09-568-472-5	Sequence 5, Appli
41	155	20.3	7257	4	US-09-567-899-5	Sequence 5, Appli
42	154	20.2	2890	4	US-09-413-814-67	Sequence 67, Appl
43	154	20.2	3798	3	US-09-335-409-6	Sequence 6, Appli
44	154	20.2	3798	4	US-09-568-102-6	Sequence 6, Appli
45	154	20.2	3798	4	US-09-567-969-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2

Query Match 50.5%; Score 385.5; DB 3; Length 3739;
Best Local Similarity 54.8%; Pred. No. 3.5e-32;
Matches 80; Conservative 14; Mismatches 47; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:|||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |         || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

```
RESULT 2
US-09-105-537-33
; Sequence 33, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-33

  Query Match             50.5%;  Score 385.5;  DB 4;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 3.5e-32;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546


RESULT 3
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             50.5%;  Score 385.5;  DB 4;  Length 11877;
  Best Local Similarity   54.8%;  Pred. No. 1.6e-31;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:||||||||| ||
Db       7355 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 7414

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       7415 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 7474

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       7475 YAHPEDAPG-----EAEWTRHATGVL 7495


RESULT 4
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match             47.8%;  Score 365.5;  DB 4;  Length 5588;
  Best Local Similarity   47.9%;  Pred. No. 8.3e-30;
  Matches   70;  Conservative   21;  Mismatches   50;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||| :  ||:: : ||   ||||||| : ||  || || |:||::: |||||| |||
Db        906 YWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQPWLADHVVGGA 965

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              :|| |  |:|: : |   |||| ::||:|  |||:|  | | ||: |   |  ||| | |
Db        966 ILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVRV 1025

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :     :|          |||||:| |
Db       1026 H-----SCRDDAVPQDSWTCHATGTL 1046


RESULT 5
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
```

```
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match             47.8%;  Score 365.5;  DB 4;  Length 5588;
  Best Local Similarity   47.9%;  Pred. No. 8.3e-30;
  Matches   70;  Conservative   21;  Mismatches   50;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||| :  ||:: : ||   ||||||| : ||  || || |:||::: |||||| |||
Db        906 YWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQPWLADHVVGGA 965

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              :|| |  |:|: : |   |||| ::||:|  |||:|  | | ||: |   |  ||| |.|
Db        966 ILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVRV 1025

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :     :|          |||||:| |
Db       1026 H-----SCRDDAVPQDSWTCHATGTL 1046


RESULT 6
US-09-036-987A-3
; Sequence 3, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2152 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-3

  Query Match             44.8%;  Score 342;  DB 4;  Length 2152;
  Best Local Similarity   47.9%;  Pred. No. 7.9e-28;
  Matches   70;  Conservative   27;  Mismatches   41;  Indels    8;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  :   :|:|: |||   ||||||| : :|     || ||:|:|:|||||:| | |
Db        920 YWLVPSD--SGDVTGAGLAGAEHPLLGAVVPVAGGDEVLLTGRISVRTHPWLAEHRVLGE 977

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              |:::|   ||:|||||  :||:||:|||| ||||:|  | : ||: | | :  ||| :::
Db        978 VIVAGTALLEIALHAGERLGCERVEELTLEAPLVLPERGAIQVQLRVGAPENSGRRPMAL 1037

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |:|         ||:   || ||:| |
Db       1038 YSR------PEGAAEHDWTRHATGRL 1057


RESULT 7
US-09-370-700-3
; Sequence 3, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2152
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-3

  Query Match             44.8%;  Score 342;  DB 4;  Length 2152;
  Best Local Similarity   47.9%;  Pred. No. 7.9e-28;
  Matches   70;  Conservative   27;  Mismatches   41;  Indels    8;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  :   :|:|: |||   ||||||| : :|     || ||:|:|:|||||:| | |
Db        920 YWLVPSD--SGDVTGAGLAGAEHPLLGAVVPVAGGDEVLLTGRISVRTHPWLAEHRVLGE 977

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              |:::|   ||:|||||  :||:||:|||| ||||:|  | : ||: | | :  ||| :::
Db        978 VIVAGTALLEIALHAGERLGCERVEELTLEAPLVLPERGAIQVQLRVGAPENSGRRPMAL 1037

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |:|         ||:   || ||:| |
Db       1038 YSR------PEGAAEHDWTRHATGRL 1057


RESULT 8
```

```
US-08-804-227C-2
; Sequence 2, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4472 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-2

  Query Match             43.7%;  Score 333.5;  DB 2;  Length 4472;
  Best Local Similarity   45.7%;  Pred. No. 1.7e-26;
  Matches   69;  Conservative   21;  Mismatches   56;  Indels    5;  Gaps    2;

Qy          1 YWLESTQPGAGNVSAA--GLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVG 58
              ||  ::  | |: :|    |:   :|| |   | ||  |  : ||||:   | || ||||
Db       3490 YWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLADSGERVFAGRLAGSEHDWLTDHAVS 3549

Qy         59 GTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLV 118
              |  || |  |:| |||||   || |::||:: ||||:|  ||| ||: |:|||  ||| |
Db       3550 GVTLLPGTAFVEFALHAGAATGCGRLEELSVEAPLVLPAAGGVRVQMRVSAADESGRRRV 3609

Qy        119 SVYARGGSA---CGGGGASGGVWTCHASGVL 146
              ::::   :|       || | |||| |  | |
Db       3610 AIHSAPEAAVHSAAEGGDSAGVWTRHGEGTL 3640


RESULT 9
US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-2

  Query Match             42.4%;  Score 324;  DB 4;  Length 5215;
  Best Local Similarity   52.0%;  Pred. No. 2.1e-25;
  Matches   77;  Conservative    9;  Mismatches   56;  Indels    6;  Gaps    4;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||  |      :  | | ||  ||||  |:|:|   | || |||||   |||||| | |
Db       1926 YWL--TPEARTDARALGFDPARHPLLTTTVEVAGGDGVLLTGRLSLTDQPWLADHMVNGA 1983

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE--GRRLV 118
              |||    |||||| || :||  ||:|||| ||||:|  | | :|||| : |||     |
Db       1984 VLLPATAFLELALAAGDHVGAVRVEELTLEAPLVLPERGAVRIQVGV-SGDGESPAGRTF 2042

Qy        119 SVYARGGSACGGGGASGGVWTCHASGVL 146
              ||:   |   | |    || | ||||
Db       2043 GVYSTPDSGDTGDDAP-REWTRHVSGVL 2069


RESULT 10
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-1

  Query Match             41.9%;  Score 320;  DB 3;  Length 4551;
  Best Local Similarity   47.1%;  Pred. No. 4.8e-25;
  Matches   74;  Conservative   19;  Mismatches   46;  Indels   18;  Gaps    4;

Qy          1 YWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              |||::  | |    : :  ||   :|||||| : |     | || |||||||:|||||||||
Db       3494 YWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRLSLRTHPWLADHAV 3551

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE---- 113
              |:||| ||  :||| ||     |   | ||||  |||:|  ||| ::| | |  ||
Db       3552 LGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELRVTVGAPAGEPGGE 3611

Qy        114 ----GRRLVSVYARGGSACGGGGASGGVWTCHASGVL 146
                  | | ||:::|   |      :|  |:|||:|:|
Db       3612 SAGDGARPVSLHSRLADA-----PAGTAWSCHATGLL 3643
```

```
RESULT 11
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             41.9%;  Score 320;  DB 4;  Length 4613;
  Best Local Similarity   47.1%;  Pred. No. 4.9e-25;
  Matches   74;  Conservative   19;  Mismatches   46;  Indels   18;  Gaps    4;

Qy          1 YWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              |||::  | |    : :  ||   :|||||| : |    | || ||||||:|||||||||
Db       3556 YWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRLSLRTHPWLADHAV 3613

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGE---- 113
               |:||| ||  :||| ||    |   | ||||  |||:|  ||| ::| | |  ||
Db       3614 LGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELRVTVGAPAGEPGGE 3673

Qy        114 ----GRRLVSVYARGGSACGGGGASGGVWTCHASGVL 146
                  | | ||:::|   |      :|  |:|||:|:|
Db       3674 SAGDGARPVSLHSRLADA-----PAGTAWSCHATGLL 3705


RESULT 12
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3

  Query Match             41.8%;  Score 319;  DB 4;  Length 3816;
  Best Local Similarity   45.6%;  Pred. No. 4.9e-25;
  Matches   67;  Conservative   27;  Mismatches   45;  Indels    8;  Gaps    3;

Qy          1 YWLE-STQPGA--GNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              || : :  |||  |:|:: ||    |||||| :|:    | :| |::|||:|||||||| |
Db       2471 YWPKPAPAPGANLGDVASVGLTAAGHPLLGAVVEMPDSDGLVLTGQISLRTHPWLADHEV 2530

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRL 117
               |:||| |  |:|||: |    | | :|||| |||:|  ||: |:: :  :: :|||
Db       2531 LGSVLLPGTAFVELAVQAADRAGYDVLDELTLEAPLVLPDRGGIQVRLALGPSEADGRRS 2590

Qy        118 VSVYARGGSACGGGGASGGVWTCHASG 144
               : :::|   | |       || ||||
Db       2591 LQLHSRPEEAAGFHR-----WTRHASG 2612


RESULT 13
US-08-804-227C-14
; Sequence 14, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-14

  Query Match             41.4%;  Score 316.5;  DB 2;  Length 4545;
  Best Local Similarity   48.3%;  Pred. No. 1.1e-24;
  Matches   73;  Conservative   14;  Mismatches   55;  Indels    9;  Gaps    3;

Qy          1 YWLESTQPGAGNVSAA---GLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              ||   :  ||   |||   ||:  :|| |     :|  |  || ||: | :|||||||:
Db       3528 YWATGSVTGATGTSAAARFGLEWKDHPFLSGATPIAGSGALLLTGRVGLAAHPWLADHAI 3587

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRL 117
               ||||| |    :| | |   ||   |:||||| ||::|  ||: ||| | ||| :|||
Db       3588 SGTVLLPGTAIADLLLRAVEEVGAGGVEELTLHEPLLLPERGGLHVQVLVEAADEQGRRA 3647

Qy        118 VSVYARGGSACGGGGASG--GVWTCHASGVL 146
               |:| ||      |.|  |    || || |||
Db       3648 VAVAARP----EGPGRDGEEQEWTRHAEGVL 3674


RESULT 14
US-08-804-227C-8
; Sequence 8, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4550 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-8

  Query Match             41.4%;  Score 316.5;  DB 2;  Length 4550;
  Best Local Similarity   48.3%;  Pred. No. 1.1e-24;
  Matches   73;  Conservative   14;  Mismatches   55;  Indels    9;  Gaps    3;

Qy          1 YWLESTQPGAGNVSAA---GLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              ||    :  ||   |||   ||:  :|| |     :|  |  || ||: | :||||||||:
Db       3533 YWATGSVTGATGTSAAARFGLEWKDHPFLSGATPIAGSGALLLTGRVGLAAHPWLADHAI 3592

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRL 117
               ||||| |    :| | |   ||   |:||||| ||::|  ||: ||| | ||| :|||
Db       3593 SGTVLLPGTAIADLLLRAVEEVGAGGVEELTLHEPLLLPERGGLHVQVLVEAADEQGRRA 3652

Qy        118 VSVYARGGSACGGGGASG--GVWTCHASGVL 146
              |:| ||      | |  |    || || |||
Db       3653 VAVAARP----EGPGRDGEEQEWTRHAEGVL 3679


RESULT 15
US-08-804-198-2
; Sequence 2, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4550 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-2

  Query Match             41.4%;  Score 316.5;  DB 2;  Length 4550;
  Best Local Similarity   48.3%;  Pred. No. 1.1e-24;
  Matches   73;  Conservative   14;  Mismatches   55;  Indels    9;  Gaps    3;

Qy          1 YWLESTQPGAGNVSAA---GLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAV 57
              ||    :  ||   |||   ||:  :|| |     :|  |  || ||: | :||||||||:
Db       3533 YWATGSVTGATGTSAAARFGLEWKDHPFLSGATPIAGSGALLLTGRVGLAAHPWLADHAI 3592

Qy         58 GGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRL 117
               ||||| |    :| | |   ||   |:||||| ||::|  ||: ||| | ||| :|||
Db       3593 SGTVLLPGTAIADLLLRAVEEVGAGGVEELTLHEPLLLPERGGLHVQVLVEAADEQGRRA 3652

Qy        118 VSVYARGGSACGGGGASG--GVWTCHASGVL 146
              |:| ||      | |  |    || || |||
Db       3653 VAVAARP----EGPGRDGEEQEWTRHAEGVL 3679


Search completed: June 17, 2003, 13:09:30
Job time : 3.65363 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 7.15215 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_2983_3128
Perfect score: 764
Sequence: 1 YWLESTQPGAGNVSAAGLDP.....ACGGGGASGGVWTCHASGVL 146

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	764	100.0	3972	21	AAB23749	S. avermitilis ave
2	764	100.0	3972	22	AAG65264	Streptomyces averm
3	764	100.0	3972	22	AAG65268	Streptomyces averm
4	753	98.6	146	21	AAB23753	S. avermitilis ave
5	585.5	76.6	5532	21	AAB23752	S. avermitilis ave
6	585.5	76.6	5532	22	AAG65267	Streptomyces averm
7	385.5	50.5	3739	21	AAB18638	Amino acid sequenc
8	385.5	50.5	3739	21	AAY77193	S. venezuelae macr
9	385.5	50.5	3739	21	AAY77201	S. venezuelae pik
10	385.5	50.5	3739	21	AAY67202	Narbonolide syntha
11	385.5	50.5	12199	21	AAY77180	S. venezuelae pik
12	372	48.7	7068	22	AAE10142	Streptomyces nours
13	372	48.7	9477	22	AAE10144	Streptomyces nours
14	365.5	47.8	5588	20	AAY39301	SpnE a polyketide
15	365.5	47.8	5588	22	AAB70969	S. spinosa protein
16	357	46.7	4881	21	AAB23751	S. avermitilis ave
17	357	46.7	4881	22	AAG65266	Streptomyces averm
18	354.5	46.4	11096	22	AAE10129	Streptomyces nours
19	351.5	46.0	6239	21	AAB23750	S. avermitilis ave
20	351.5	46.0	6239	22	AAG65265	Streptomyces averm
21	348	45.5	6797	22	AAB31558	Pimaricin biosynth
22	342	44.8	2152	20	AAY39298	SpnB a polyketide
23	342	44.8	2152	22	AAB70966	S. spinosa protein
24	333.5	43.7	4472	18	AAW22601	Tylactone synthase
25	332	43.5	5069	19	AAW52846	A. mediterranei ri
26	324	42.4	4630	18	AAW19629	Streptomyces venez
27	324	42.4	4630	21	AAY77177	S. venezuelae vep
28	320	41.9	4551	21	AAB18637	Amino acid sequenc
29	320	41.9	4551	21	AAY67201	Narbonolide syntha
30	320	41.9	4613	21	AAY77192	S. venezuelae macr
31	320	41.9	4613	21	AAY77200	S. venezuelae pik
32	319	41.8	3816	21	AAY92708	S. antibioticus 8,
33	316.5	41.4	4545	18	AAW22611	Hybrid srmG/tylG O
34	316.5	41.4	4550	18	AAW23716	Platenolide syntha
35	316.5	41.4	4550	18	AAW22606	Platenolide syntha
36	311.5	40.8	4924	22	AAB70968	S. spinosa protein
37	311.5	40.8	4928	20	AAY39300	SpnD a polyketide
38	309.5	40.5	5435	22	AAE10145	Streptomyces nours
39	308.5	40.4	1891	19	AAW52844	Amycolatopsis medi
40	308.5	40.4	3413	19	AAW52849	A. mediterranei ri
41	308.5	40.4	3729	18	AAW22603	Tylactone synthase
42	304	39.8	1615	22	AAG81211	Mycobacterium tube
43	300	39.3	1721	19	AAW52847	A. mediterranei ri
44	291	38.1	5087	22	AAU10700	Amino acid sequenc
45	279.5	36.6	1366	22	AAE10127	Streptomyces nours

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 764; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2e-68;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
Db 2983 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 3042

Qy 61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
Db 3043 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 3102

Qy 121 YARGGSACGGGGASGGVWTCHASGVL 146
Db 3103 YARGGSACGGGGASGGVWTCHASGVL 3128

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 764; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2e-68;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
Db 2983 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 3042

Qy 61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
Db 3043 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 3102

Qy 121 YARGGSACGGGGASGGVWTCHASGVL 146
Db 3103 YARGGSACGGGGASGGVWTCHASGVL 3128

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.

```
XX
SQ   Sequence   3972 AA;

  Query Match            100.0%;  Score 764;  DB 22;  Length 3972;
  Best Local Similarity  100.0%;  Pred. No. 2e-68;
  Matches  146;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   .   2983 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 3042

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3043 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 3102

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              |||||||||||||||||||||||||
Db      3103 YARGGSACGGGGASGGVWTCHASGVL 3128


RESULT 4
AAB23753
ID   AAB23753 standard; Protein; 146 AA.
XX
AC   AAB23753;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:8.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92303.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Disclosure; Page 307-308; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   146 AA;

  Query Match            98.6%;  Score 753;  DB 21;  Length 146;
  Best Local Similarity  98.6%;  Pred. No. 4.4e-69;
  Matches  144;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADYEVGGT 60

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||||||||||||||||||||||||||
Db        121 YARGGSACGGGGASGGVWTCHASGVL 146


RESULT 5
AAB23752
ID   AAB23752 standard; Protein; 5532 AA.
XX
AC   AAB23752;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 278-305; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;

  Query Match            76.6%;  Score 585.5;  DB 21;  Length 5532;
  Best Local Similarity  76.0%;  Pred. No. 5.1e-50;
  Matches  111;  Conservative   15;  Mismatches   19;  Indels    1;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              |||::   |||:|:||||:| ||||| ||::||   | || ||||||||||| |: |||
Db        955 YWLDA-PTGAGDVAAAGLEPAEHPLLAATVQLADTDGCLLTGRLSLRSHPWLGDYEVGGA 1013
```

```
Qy          61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               |||||: |:|||:  |  ||| |:::||:|||||||| ||||||||||||||||||||
Db        1014 VLLSGSAFVELAVQVGERVGCTRIEQLTVHAPLVVPVGGGVSVQVGVAAADGEGRRLVSV 1073

Qy         121 YARGGSACGGGGASGGVWTCHASGVL 146
               |||||||||||||||||||||||||
Db        1074 YARGGSACGGGGASGGVWTCHASGVL 1099
```

RESULT 6
AAG65267
ID AAG65267 standard; Protein; 5532 AA.
XX
AC AAG65267;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 7.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79278.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 217-235; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 5532 AA;

Query Match 76.6%; Score 585.5; DB 22; Length 5532;
Best Local Similarity 76.0%; Pred. No. 5.1e-50;
Matches 111; Conservative 15; Mismatches 19; Indels 1; Gaps 1;

```
Qy           1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
               |||::   |||:|:||||:| ||||| ||::||   | || ||||||||||| |: |||
Db         955 YWLDA-PTGAGDVAAAGLEPAEHPLLAATVQLADTDGCLLTGRLSLRSHPWLGDYEVGGA 1013

Qy          61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               |||||: |:|||:  |  ||| |:::||:|||||||| ||||||||||||||||||||
Db        1014 VLLSGSAFVELAVQVGERVGCTRIEQLTVHAPLVVPVGGGVSVQVGVAAADGEGRRLVSV 1073

Qy         121 YARGGSACGGGGASGGVWTCHASGVL 146
               |||||||||||||||||||||||||
Db        1074 YARGGSACGGGGASGGVWTCHASGVL 1099
```

RESULT 7
AAB18638
ID AAB18638 standard; Protein; 3739 AA.
XX
AC AAB18638;
XX
DT 22-JAN-2001 (first entry)
XX
DE Amino acid sequence of narbonolide synthase subunit 2 (PICAII).
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 11-12; 117pp; English.
XX
CC The present sequence represents a narbonolide synthase subunit 2
CC (PICAII). The nucleotide sequence encoding it is used in the course of
CC the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 3739 AA;

Query Match 50.5%; Score 385.5; DB 21; Length 3739;
Best Local Similarity 54.8%; Pred. No. 8.3e-30;
Matches 80; Conservative 14; Mismatches 47; Indels 5; Gaps 1;

```
Qy           1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
               || :     ||::::|||   ||||||| : ||   | || | ||||:|||||||| ||
Db        2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy          61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
```

```
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

RESULT 8
AAY77193
ID AAY77193 standard; Protein; 3739 AA.
XX
AC AAY77193;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae macrolide biosynthetic enzyme PikAII, SEQ ID NO:33.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87298.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 19; Page 403-415; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC AAZ87295-Z87302.
XX
SQ Sequence 3739 AA;

```
  Query Match             50.5%;  Score 385.5;  DB 21;  Length 3739;
  Best Local Similarity   54.8%;  Pred. No. 8.3e-30;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:|||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

RESULT 9
AAY77201
ID AAY77201 standard; Protein; 3739 AA.
XX
AC AAY77201;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster protein #2.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87318.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and

CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77200-Y77203 and AAY80997 represent proteins encoded by the
CC macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
CC ATCC 15439, as given in figure 31.
XX
SQ Sequence 3739 AA;

Query Match 50.5%; Score 385.5; DB 21; Length 3739;
Best Local Similarity 54.8%; Pred. No. 8.3e-30;
Matches 80; Conservative 14; Mismatches 47; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:|||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

RESULT 10
AAY67202
ID AAY67202 standard; protein; 3739 AA.
XX
AC AAY67202;
XX
DT 23-MAR-2000 (first entry)
XX
DE Narbonolide synthase subunit 2 (PICAII) protein sequence.
XX
KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 2;
KW PICAII; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS Streptomyces venezuelae.
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR N-PSDB; AAZ56001.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Example 2; Page 13-14; 98pp; English.
XX
CC This is the Streptomyces venezuelae narbonolide synthase subunit 2,
CC PICAII protein sequence. The invention relates to recombinant DNA
CC containing a coding sequence for a narbonolide polyketide synthase
CC (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC a series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC transferase enzyme is required for this conversion, and the desosamine
CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC DNA of the invention is used to express, in transformed cells,
CC narbonolide (or its derivatives) or other ketolides (particularly
CC hybrids), which may then be converted (e.g. by other enzymes
CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.
XX
SQ Sequence 3739 AA;

Query Match 50.5%; Score 385.5; DB 21; Length 3739;
Best Local Similarity 54.8%; Pred. No. 8.3e-30;
Matches 80; Conservative 14; Mismatches 47; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :     ||::::|||   ||||||| : ||   | || | ||||:|||||||||| ||
Db       2406 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 2465

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       2466 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 2525

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       2526 YAHPEDAPG-----EAEWTRHATGVL 2546
```

RESULT 11
AAY77180
ID AAY77180 standard; Protein; 12199 AA.
XX
AC AAY77180;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX

```
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87285.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 315-353; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents a protein
CC   encoded by the macrolide biosynthetic gene cluster (pik) from
CC   Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence   12199 AA;

  Query Match             50.5%;  Score 385.5;  DB 21;  Length 12199;
  Best Local Similarity   54.8%;  Pred. No. 3.6e-29;
  Matches   80;  Conservative   14;  Mismatches   47;  Indels    5;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||  :     ||::::|||   ||||||| : ||   | || | ||||:|||||||||| ||
Db       7677 YWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGT 7736

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              ||| |  |:|||  ||  |||| |:|||| ||||:|  | | ||: | |:|  |||   :
Db       7737 VLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGL 7796

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              ||    | |        || ||:|||
Db       7797 YAHPEDAPG-----EAEWTRHATGVL 7817


RESULT 12
AAE10142
ID   AAE10142 standard; Protein; 7068 AA.
XX
AC   AAE10142;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (partial).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          34..448
FT                   /label= KS9_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          572..890
FT                   /label= AT9_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          904..1123
FT                   /label= DH9_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1443..1686
FT                   /label= KR9_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1720..1793
FT                   /label= ACP9_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1813..2236
FT                   /label= KS10_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2346..2664
FT                   /label= AT10_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2678..2890
FT                   /label= DH10_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          2983..3229
FT                   /label= KR10_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3266..3339
FT                   /label= ACP10_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3358..3780
FT                   /label= KS11_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          3898..4217
FT                   /label= AT11_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4231..4432
FT                   /label= DH11_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4523..4770
FT                   /label= KR11_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          4806..4879
FT                   /label= ACP11_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          4801..5325
FT                   /label= KS12_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          5432..5754
FT                   /label= AT12_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          5768..5977
FT                   /label= DH12_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          6068..6315
FT                   /label= KR12_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          6348..6421
FT                   /label= ACP12_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          6454..6873
FT                   /label= KS13_domain
FT                   /note= "Ketosynthase (KS) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
```

```
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17185.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 181-185; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (partial).
XX
SQ   Sequence   7068 AA;

  Query Match             48.7%;  Score 372;  DB 22;  Length 7068;
  Best Local Similarity   51.4%;  Pred. No. 4.4e-28;
  Matches   75;  Conservative   18;  Mismatches   47;  Indels    6;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :: | |  :: : ||    |||| | : ||  || || |||| ::|||||| | ||
Db        914 YWPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRGT 973

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               || |  |||||: ||  ||||||:|||| |||::|  ||| ||: :   |  ||| |:|
Db        974 TLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRTVNV 1033

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :||  :           || ||:|||
Db       1034 HARPDT------GDDTPWTAHATGVL 1053


RESULT 13
AAE10144
ID   AAE10144 standard; Protein; 9477 AA.
XX
AC   AAE10144;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (complete).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          41..464
FT                   /label= KS15_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          578..889
FT                   /label= AT15_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          903..1102
FT                   /label= DH15_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1446..1731
FT                   /label= ER15_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          1740..1988
FT                   /label= KR15_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          2023..2096
FT                   /label= ACP15_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          2117..2538
FT                   /label= KS16_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2635..2953
FT                   /label= AT16_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2967..3167
FT                   /label= DH16_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3257..3500
FT                   /label= KR16_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3539..3612
FT                   /label= ACP16_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3634..4057
FT                   /label= KS17_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4153..4472
FT                   /label= AT17_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4486..4725
FT                   /label= DH17_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4997..5245
FT                   /label= KR17_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5277..5350
FT                   /label= ACP17_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 255-260; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
```

```
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (complete).
XX
SQ   Sequence   9477 AA;

  Query Match             48.7%;  Score 372;  DB 22;  Length 9477;
  Best Local Similarity   51.4%;  Pred. No. 6.3e-28;
  Matches   75;  Conservative   18;  Mismatches   47;  Indels    6;  Gaps    1;

Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              || :: | |  :: : ||    |||| | : ||  || || |||| ::|||||| | ||
Db        914 YWPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRGT 973

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
               || |  |||||: ||  ||||||:|||| |||::|  ||| ||: :  '  |  ||| |:|
Db        974 TLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRTVNV 1033

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :||  :           || ||:|||
Db       1034 HARPDT------GDDTPWTAHATGVL 1053


RESULT 14
AAY39301
ID   AAY39301 standard; Protein; 5588 AA.
XX
AC   AAY39301;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnE a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..424
FT                   /label= KS8
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 8"
FT   Domain          530..848
FT                   /label= AT8
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 8"
FT   Domain          883..1070
FT                   /label= DH8
FT                   /note= "Dehydratase domain: part of extender
FT                         module 8"
FT   Domain          1369..1552
FT                   /label= KR8
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 8"
FT   Domain          1648..1726
FT                   /label= ACP8
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 8"
FT   Domain          1749..2173
FT                   /label= KS9
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 9"
FT   Domain          2286..2614
FT                   /label= AT9
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 9"
FT   Domain          2640..2800
FT                   /label= DH9
FT                   /note= "Dehydratase domain: part of extender
FT                         module 9"
FT   Domain          3157..3341
FT                   /label= KR9
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 9"
FT   Domain          3422..3500
FT                   /label= ACP9
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 9"
FT   Domain          3534..3948
FT                   /label= KS10
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 10"
FT   Domain          4060..4390
FT                   /label= AT10
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 10"
FT   Domain          4413..4597
FT                   /label= DH10
FT                   /note= "Dehydratase domain: part of extender
FT                         module 10"
FT   Domain          4900..5078
FT                   /label= KR10
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 10"
FT   Domain          5172..5253
FT                   /label= ACP10
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 10"
FT   Domain          5302..5555
FT                   /label= TE10
FT                   /note= "Thiolester domain: part of extender
FT                         module 10"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 130-148; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnE gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnE gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
```

CC several domains with specific functions. The initiator module has an
CC acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC The extender modules have the same domains plus a beta-ketosynthase (KS)
CC domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC domain, and an enoyl reductase (ER) domain. The last extender module
CC terminates with a thiolester domain. The products of the genes present
CC in the upstream region the PKS genes have been assigned names spnF-spnS
CC AAY39302-Y39315 and are responsible for different modifications in
CC spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC or interruption of steps in spinosyn biosynthesis. The modified
CC spinosyns may be a new insect control agent or serve as substrates for
CC further chemical modification and the creation of new semi-synthetic
CC spinosyns. The genes are also useful to isolate similar sequences from S.
CC spinosa or other species by hybridization.
XX
SQ Sequence 5588 AA;

Query Match 47.8%; Score 365.5; DB 20; Length 5588;
Best Local Similarity 47.9%; Pred. No. 1.5e-27;
Matches 70; Conservative 21; Mismatches 50; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||| :  ||:: : ||   ||||||| : ||  || || |:||::: |||||| |||
Db        906 YWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQPWLADHVVGGA 965

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              :|| |  |:|: : |   |||| ::||:|  |||:|  | | ||: |   |  ||| | |
Db        966 ILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVRV 1025

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :    :|         |||||:| |
Db       1026 H-----SCRDDAVPQDSWTCHATGTL 1046
```

RESULT 15
AAB70969
ID AAB70969 standard; Protein; 5588 AA.
XX
AC AAB70969;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa protein fragment encoded by ORF22, SEQ ID 50.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR N-PSDB; AAF88339.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 57; Page 313-335; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ Sequence 5588 AA;

Query Match 47.8%; Score 365.5; DB 22; Length 5588;
Best Local Similarity 47.9%; Pred. No. 1.5e-27;
Matches 70; Conservative 21; Mismatches 50; Indels 5; Gaps 1;

```
Qy          1 YWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGT 60
              ||||| :  ||:: : ||   ||||||| : ||  || || |:||::: |||||| |||
Db        906 YWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQPWLADHVVGGA 965

Qy         61 VLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSV 120
              :|| |  |:|: : |   |||| ::||:|  |||:|  | | ||: |   |  ||| | |
Db        966 ILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVRV 1025

Qy        121 YARGGSACGGGGASGGVWTCHASGVL 146
              :    :|         |||||:| |
Db       1026 H-----SCRDDAVPQDSWTCHATGTL 1046
```

Search completed: June 17, 2003, 13:02:33
Job time : 8.15215 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 19.8736 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_2086_2515
Perfect score: 2265
Sequence: 1 EPIAIVGMACRFPGGVTSAD.....VSSFGVSGTNAHVILEEAPA 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2265	100.0	3626	2	Q9F7T9	Q9f7t9 streptomyce
2	2265	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
3	2042	90.2	6239	2	Q9S0R7	Q9s0r7 streptomyce
4	1968.5	86.9	5532	2	Q9S0R4	Q9s0r4 streptomyce
5	1744.5	77.0	4881	2	Q9S0R3	Q9s0r3 streptomyce
6	1733.5	76.5	6048	2	Q93H87	Q93h87 streptomyce
7	1704	75.2	3170	2	Q9ALM4	Q9alm4 saccharopol
8	1700.5	75.1	6145	2	Q93H84	Q93h84 streptomyce
9	1672	73.8	4928	2	Q9ALM3	Q9alm3 saccharopol
10	1668.5	73.7	3524	2	Q93H86	Q93h86 streptomyce
11	1656.5	73.1	3352	2	Q93H83	Q93h83 streptomyce
12	1656	73.1	11096	2	Q9L4W3	Q914w3 streptomyce
13	1646.5	72.7	4150	2	Q9KIV4	Q9kiv4 streptomyce
14	1645.5	72.6	6797	2	Q9X993	Q9x993 streptomyce
15	1641	72.5	10917	2	Q93NW6	Q93nw6 streptomyce
16	1640	72.4	5588	2	Q9ALM2	Q9alm2 saccharopol
17	1638	72.3	2723	2	O30479	O30479 streptomyce
18	1636	72.2	5060	2	O52545	O52545 amycolatops
19	1636	72.2	5069	2	O52789	O52789 amycolatops
20	1634.5	72.2	1835	2	Q93H85	Q93h85 streptomyce
21	1632	72.1	6146	2	Q93HJ5	Q93hj5 streptomyce
22	1631.5	72.0	4472	2	O33954	O33954 streptomyce
23	1628.5	71.9	3739	2	Q9ZGI4	Q9zgi4 streptomyce
24	1626.5	71.8	3970	2	Q93HI8	Q93hi8 streptomyce
25	1625.5	71.8	4340	2	O30764	O30764 streptomyce
26	1622.5	71.6	2595	2	Q9ALM6	Q9alm6 saccharopol
27	1622.5	71.6	4613	2	Q9ZGI5	Q9zgi5 streptomyce
28	1619.5	71.5	2152	2	Q9ALM5	Q9alm5 saccharopol
29	1614.5	71.3	9477	2	Q9L4X3	Q914x3 streptomyce
30	1613.5	71.2	3576	16	Q9EX53	Q9ex53 streptomyce
31	1613.5	71.2	4498	2	Q93HE5	Q93he5 streptomyce
32	1611	71.1	4735	2	O54666	O54666 amycolatops
33	1609	71.0	3816	2	Q9KIV3	Q9kiv3 streptomyce
34	1608.5	71.0	9510	2	Q93NX9	Q93nx9 streptomyce
35	1606	70.9	1265	2	Q9F847	Q9f847 amycolatops
36	1605	70.9	7525	2	Q9KIE0	Q9kie0 streptomyce
37	1604.5	70.8	3939	2	Q93HJ3	Q93hj3 streptomyce
38	1601.5	70.7	9507	2	Q9EWA1	Q9ewa1 streptomyce
39	1597	70.5	3729	2	O33956	O33956 streptomyce
40	1594.5	70.4	1839	2	O30765	O30765 streptomyce
41	1594.5	70.4	2066	2	Q9L4X1	Q914x1 streptomyce
42	1593.5	70.4	5435	2	Q9L4X2	Q914x2 streptomyce
43	1591.5	70.3	3413	2	O54593	O54593 amycolatops
44	1590	70.2	7576	2	Q9ZGA4	Q9zga4 streptomyce
45	1589.5	70.2	3654	2	O30766	O30766 streptomyce

ALIGNMENTS

RESULT 1
Q9F7T9
ID Q9F7T9 PRELIMINARY; PRT; 3626 AA.
AC Q9F7T9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Avermectin polyketide synthase (Fragment).
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31271;
RA Hong Y.-S., Lee J.J.;
RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT and polyketide synthase gene from Streptomyces avermitilis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275943; AAG09812.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT NON_TER 3626 3626
SQ SEQUENCE 3626 AA; 380557 MW; 6272F5F088C1A8D0 CRC64;

Query Match 100.0%; Score 2265; DB 2; Length 3626;
Best Local Similarity 100.0%; Pred. No. 3.2e-119;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60

```
Db       2053 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2112

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2113 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2172

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2173 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 2232

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2233 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 2292

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2293 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 2352

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2353 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 2412

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2413 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 2472

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       2473 AHVILEEAPA 2482
```

```
RESULT 2
Q9S0R8
ID   Q9S0R8      PRELIMINARY;      PRT;  3972 AA.
AC   Q9S0R8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 1.
GN   AVEA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84474.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3972 AA;  416852 MW;  2A293695B032B1C3 CRC64;

  Query Match           100.0%;  Score 2265;  DB 2;  Length 3972;
  Best Local Similarity  100.0%;  Pred. No. 3.6e-119;
  Matches  430;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2086 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2145

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2146 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2205

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2206 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 2265

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2266 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 2325

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2326 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 2385

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2386 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 2445

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2446 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 2505

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       2506 AHVILEEAPA 2515
```

```
RESULT 3
Q9S0R7
ID   Q9S0R7      PRELIMINARY;      PRT;  6239 AA.
AC   Q9S0R7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 2.
GN   AVEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84475.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6239 AA;  666283 MW;  A706280940B502FA CRC64;

  Query Match           90.2%;  Score 2042;  DB 2;  Length 6239;
  Best Local Similarity  89.5%;  Pred. No. 2.4e-106;
  Matches  385;  Conservative   14;  Mismatches   31;  Indels    0;  Gaps    0;
```

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2680 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2739

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2740 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2799

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |||||||:|  :  |   |:|| ||| | |: ||||||  | ||||||::||||||||||
Db       2800 TGTNGQDHAAHIRQAPSGTEGFVLTGAATSIASGRISYILGLEGPAVTLDTACSSSLVAL 2859

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:||:|||:|||||| |||::|  | ||:|||||| || ||||||||||||||||||
Db       2860 HLACQSLRSGECTMALAGGATVMTTPITFTEFARQRGLAPDGRCKAFSAAADGTGWGEGV 2919

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||| |:
Db       2920 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANADLTPA 2979

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||||||||||:|||||||| | ||||||||||||||||||||||||||||||
Db       2980 DVDAVEAHGTGTTLGDPIEAQAILATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIK 3039

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              |||||||  || ||| ||||||||||||||||||||||||||||| ||||||||||||||
Db       3040 MVMALRHRTLPPTLHADEPSPHVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTN 3099

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       3100 AHVILEEAPA 3109
```

RESULT 4
Q9S0R4
ID Q9S0R4 PRELIMINARY; PRT; 5532 AA.
AC Q9S0R4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 3.
GN AVEA3.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84478.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 5532 AA; 575193 MW; EC7B2069D44A4057 CRC64;

Query Match 86.9%; Score 1968.5; DB 2; Length 5532;
Best Local Similarity 87.2%; Pred. No. 2.9e-102;
Matches 376; Conservative 23; Mismatches 31; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||:|||||:|||||||:: |:|::| :| :| |||||||||: |:||:|  ||| ||
Db       1864 EPVAIIGMACRYPGGVTSAEELWELLASGRDTVGEFPTDRGWDLEALFDPEPGRPGTSYT 1923

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | |||||| ||| |||||||||||||||||||||| :|| :| |||:| |: |: ||||
Db       1924 RCGSFLYDAGEFDAGFFGISPREALAMDPQQRLLLEASWEAMEQAGIDPTTVRGSQTGVF 1983

Qy        121 TGTNGQDYALRVH-NAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179
              |  | | |:| ||   |:|: ||||:||| ||||||||||||||||||||||||||||
Db       1984 AGLIPQAYGPRLHENAAADTEGYVLTGTSGSVASGRISYTFGFEGPAVSVDTACSSSLVA 2043

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2044 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 2103

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2104 VGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 2163

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2164 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 2223

Qy        360 KMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 419
              |||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2224 KMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 2283

Qy        420 NAHVILEEAPA 430
              ||||||||||
Db       2284 NAHVILEEAPA 2294
```

RESULT 5
Q9S0R3
ID Q9S0R3 PRELIMINARY; PRT; 4881 AA.
AC Q9S0R3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 4.
GN AVEA4.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84479.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 4881 AA; 510303 MW; F6568C1F01ACD37A CRC64;

Query Match 77.0%; Score 1744.5; DB 2; Length 4881;
Best Local Similarity 76.7%; Pred. No. 1.1e-89;
Matches 329; Conservative 40; Mismatches 59; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||||||| ||:|||:||:| :||:| || |||||::  |||:|    || ||
Db       2636 EPIAIVGMACRFPGGVESAEDFWELIASGRDAVGEFPVDRGWDVEAFYDPEPGRAGTSYT 2695

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | ||||  |  ||| ||||||||||||||||||:|| :|| :| |||:| ||||: ||||
Db       2696 RCGGFLQGAAEFDAGFFGISPREALAMDPQQRLMLEVSWEALERAGIDPATLHGSTTGVF 2755

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              | : ||||  :    |  :|:|||| : ||:|||:||||||||||:||||||||||||
Db       2756 AGVSQQDYAELLRRGTQDHEGYALTGVSNSVVSGRLSYTFGFEGPAVTVDTACSSSLVAL 2815

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||:||||:||||||||||:|||||||||||||: || |||: : ||| || |||
Db       2816 HLACQALRSGECSLALAGGVTVMSTPGAFVEFSRQRGLSPDGRCKAYGSGADGVGWAEGV 2875

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||:| | ||||||||||||||||||||||||||||||||||||||||| ||||
Db       2876 GVLLVERLSEAERRGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALACAGLSVA 2935

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| || |||||||||||||||||||||| |:|| |:|||||||:|| ||||||||||
Db       2936 DVDVVEGHGTGTTLGDPIEAQALLATYGQGRSGERPVWLGSVKSNIGHAQAAAGVAGVIK 2995

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||| | ||| :||:||||||:|||:| |:|||| |||     || ||||||:|||||||
Db       2996 MVMALNHELLPTSLHIDEPSPHIDWSSGGVRLLTEPVPWQ-QNGRPRRAGVSAFGVSGTN 3054

Qy        421 AHVILEEAP 429
              ||||:|:||
Db       3055 AHVIIEQAP 3063
```

RESULT 6
Q93H87
ID Q93H87 PRELIMINARY; PRT; 6048 AA.
AC Q93H87;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Modular polyketide synthase.
GN PTEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMBL; AB070949; BAB69303.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 4.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR TIGRFAMs; TIGR00128; fabD; 4.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
KW Phosphopantetheine.
SQ SEQUENCE 6048 AA; 629420 MW; F8673C2ED2694705 CRC64;

Query Match 76.5%; Score 1733.5; DB 2; Length 6048;
Best Local Similarity 76.3%; Pred. No. 5.8e-89;
Matches 328; Conservative 42; Mismatches 59; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||:|| |||| | :|  |:|:|| :||: |||||||||:| ||||||| ||  |:
Db       2552 EPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDIDALYDPDPDRPGKTYS 2611

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||:  |  |||  ||||||||||||||||||||||  | |||   ::  : ||||
Db       2612 REGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFERAGIALASVRASRTGVF 2671

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||||||  : || :  :|:|||| | ||:|||||||| |||||:||||||||||||
Db       2672 IGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFGLEGPAVTVDTACSSSLVAL 2731

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||||:|||||| |||||||:|  ||||||||||: || ||||:| ||||||||||
Db       2732 HLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLSPDGRCKAFAAGADGTGWGEGV 2791

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||:|||:|||||||||||||||||||||||||||||:|||:
Db       2792 GLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASAGLTPA 2851

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| |||||||||||||||: : : |:|||||||:|||||||||||||
Db       2852 DVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDRPVWLGSVKSNIGHTQAAAGVAGVIK 2911

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||:|:|||:||:|||||||: |||||||||:|||| | |||   |  ||| |||||: |||
Db       2912 MVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLLTEPVQWPGTTGP-RRAAVSSFGIGGTN 2970

Qy        421 AHVILEEAPA 430
              || |:|||||
Db       2971 AHTIIEEAPA 2980
```

RESULT 7
Q9ALM4
ID Q9ALM4 PRELIMINARY; PRT; 3170 AA.
AC Q9ALM4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Polyketide synthase extender modules 3-4.
GN SPNC.
OS Saccharopolyspora spinosa.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=60894;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21257765; PubMed=11358695;
RA Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT "Cloning and analysis of the spinosad biosynthetic cluster of
RT Saccharopolyspora spinosa.";
RL Chem. Biol. 8:487-499(2001).
DR EMBL; AY007564; AAG23266.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 2.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.

```
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3170 AA;  334873 MW;  205D519FDFA3249C CRC64;

  Query Match            75.2%;  Score 1704;  DB 2;  Length 3170;
  Best Local Similarity  75.3%;  Pred. No. 1.2e-87;
  Matches  323;  Conservative   38;  Mismatches   66;  Indels    2;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| | ||:|||| : | |::| |||| || |||| || ||||||| ||||||
Db         34 EPIAIVAMGCRYPGGVQDPEGLWKLVASGGDAIGEFPADRGWHLDELYDPDPDQPGTCYT 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| ||| || |||||||||||||||||| :|||:| ||::| :| |: ||||
Db         94 RHGGFLHDAGEFDAGFFDISPREALAMDPQQRLLLEISWETVESAGMDPRSLRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |   : |    | ||:  :|:  || |||| |||::| |||||||:||||||||||
Db        154 AGLMYEGYDTGAHRAGEGVEGYLGTGNAGSVASGRVAYAFGFEGPAVTVDTACSSSLVAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:|| ||| :|||||||||:|  ||||||||||| || ||:|:||||||||||||
Db        214 HLACQSLRQGECDLALAGGVTVMSTPERFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGA 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |::|:|||||| |||||||||||||||||||||||||||| :|:|||:| | :|||||
Db        274 GLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGLAQERVIQQVLTSAGLSAS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| ||||||||:| |||||  : ||||||||||:||||||||||||||
Db        334 DVDAVEAHGTGTRLGDPIEAQALIAAYGQDRDRDRPLWLGSVKSNIGHTQAAAGVAGVIK 393

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:||| |||||||||: |||||||||:|||| : |||      ||||||||||||||
Db        394 MVMAMRHGELPRTLHVDEPNSHVDWSAGAVRLLTENIRWPG--TGTRRAGVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              |||||| |
Db        452 AHVILEHDP 460


RESULT 8
Q93H84
ID   Q93H84      PRELIMINARY;      PRT;  6145 AA.
AC   Q93H84;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69306.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_4.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
KW   Phosphopantetheine.
SQ   SEQUENCE   6145 AA;  642013 MW;  6A75B3EF7A7EDE95 CRC64;

  Query Match            75.1%;  Score 1700.5;  DB 2;  Length 6145;
  Best Local Similarity  75.5%;  Pred. No. 4.3e-87;
  Matches  324;  Conservative   39;  Mismatches   65;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||:|||| | :| | |:::  ||:| || |||||:: :|||||  | |
Db       4576 EPIAIVGMACRYPGGVRSPEDLWHLVAAGGDAVGDFPADRGWDVEDIYDPDPDARGKTYA 4635

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||||:|| ||  |||||||||:|||||||||||| ||| | |||:  :: |: ||||
Db       4636 RRGGFLYEAGEFDPGFFGISPREAVAMDPQQRLLLETTWETFERAGIDAESVRGSRTGVF 4695

Qy    *   121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |:  |||| |  ||   ::||  ||.: |::||||:|||| |||||:|||||||||||
Db       4696 VGSGYQDYAARAFNAIDDSEGFFGTGNSASIMSGRIAYTFGLEGPAVTVDTACSSSLVAL 4755

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              | | |||| |||||||||| |||:| |||||||||| || |||| | |||||| |||
Db       4756 HWAIQALRNGECSMALAGGVMVMSTPRAFVEFSRQRGLAPDGRCKAFGADADGTGWAEGV 4815

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||| ||||:|||||||||||||||||||||||:||||||||| :|||:|
Db       4816 GMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALTSAGLTAA 4875

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |:|||||||||| |||||||||||||:|    ||||||:|||:|||||||||||||
Db       4876 DIDAVEAHGTGTRLGDPIEAQALLATYGQERPAGQPLWLGSLKSNIGHTQAAAGVAGVIK 4935

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:|||:||:|||||||:|||||||| |:|||| | ||  | | ||| ||||||||||
Db       4936 MVMAMRHGVLPQTLHVDEPTPHVDWSAGDVRLLTEAVAWPESE-RPRRAAVSSFGVSGTN 4994

Qy        421 AHVILEEAP 429
              || |:|:||
Db       4995 AHTIIEQAP 5003


RESULT 9
Q9ALM3
ID   Q9ALM3      PRELIMINARY;      PRT;  4928 AA.
AC   Q9ALM3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase extender modules 5-7.
GN   SPND.
OS   Saccharopolyspora spinosa.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23263.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
```

```
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4928 AA;  517247 MW;  C22C6361888F865A CRC64;

  Query Match             73.8%;  Score 1672;  DB 2;  Length 4928;
  Best Local Similarity   75.2%;  Pred. No. 1.3e-85;
  Matches  322;  Conservative   34;  Mismatches   70;  Indels    2;  Gaps    2;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||||| |||||||:| :: | |::   ||: ||| ||||||  |||||||  || |
Db       1787 VAIVGMGCRFPGGVSSPEELWRLVAGGVDAVAGFPDDRGWDLAALYDPDPDRLGTSYVCE 1846

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTG 122
              |||| ||  |||: |||||||||||||||||||| ||||:| |||:| :|||: |||| |
Db       1847 GGFLRDAAEFDADMFGISPREALAMDPQQRLLLEVAWETLERAGIDPFSLHGSRTGVFAG 1906

Qy        123 TNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
                 ||  | :  | :  :|   || ||||:|||::|:||||||||:|||||||||||||
Db       1907 LMYHDYGARFITRAPEGFEGHLGTGNAGSVLSGRVAYSFGFEGPAVTVDTACSSSLVALH 1966

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              || |||||||  ||||||||||:|  |||||||||| || ||:|:|||||||||| |
Db       1967 LAGQALRAGECEFALAGGVTVMSTPTTFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGAG 2026

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
              ::|:|||||| |||| |||||||||||||||||||||||||||| ||| :|||| |
Db       2027 LVLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALTSAGLSVSD 2086

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              |||||||||| ||||||||||:||||:||    ||||||||||:|||||||||||||||
Db       2087 VDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAGVIKM 2146

Qy        362 VMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNA 421
              |||:| | |||||||||| |||||| |||||| |||   ||||||||||||:|||||
Db       2147 VMAMRQGELPRTLHVDEPSAQVDWSAGTVQLLTENTPWP-DSGRLRRAGVSSFGISGTNA 2205

Qy        422 HVILEEAP 429
              |:|||: |
Db       2206 HLILEQPP 2213


RESULT 10
Q93H86
ID   Q93H86      PRELIMINARY;      PRT;  3524 AA.
AC   Q93H86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69304.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3524 AA;  366619 MW;  64A65759C461EBC4 CRC64;

  Query Match             73.7%;  Score 1668.5;  DB 2;  Length 3524;
  Best Local Similarity   74.1%;  Pred. No. 1.4e-85;
  Matches  318;  Conservative   40;  Mismatches   70;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||:|| |||| | :| | |::| :| |  || :||||:| ||||||:  |  |
Db       1730 EPIAIVGMSCRLPGGVRSPEDLWRLVASGRDGITRFPDNRGWDVDGLYDPDPERLGKTYA 1789

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||::|  ||| |||||||||||||||||||| :||  | |||:| |: |: ||||
Db       1790 RDGGFLHEAAEFDAGFFGISPREALAMDPQQRLLLEASWEAFERAGIDPATIRGSRTGVF 1849

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    ||        :  :|   || :||| |||:||||| |||||::||||||||||
Db       1850 TGLMYHDYGQGPRQLPEGVEGLLTTGGSGSVASGRVSYTFGLEGPAVTIDTACSSSLVAL 1909

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |:|| |||:|||||||||: |  ||||||||||: || ||||:| ||||||||||
Db       1910 HLAVQSLRNGECTMALAGGVTVMAGPTVFVEFSRQRGLSPDGRCKAFAAGADGTGWGEGV 1969

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||||||||||||||||||||||||||:||||||||||:|||
Db       1970 GLLLVERLSDAERNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALASAGLKPA 2029

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||||||||||||||||||||:|    ||||||:|||:|||||||||||||
Db       2030 DVDAVEAHGTGTTLGDPIEAQALLATYGQERPAGQPLWLGSLKSNIGHTQAAAGVAGVIK 2089

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:|||:||:|||||||:|||||||| |:|||| | ||  : | ||| ||||||||||
Db       2090 MVMAMRHGVLPQTLHVDEPTPHVDWSAGDVRLLTEAVAWPDSD-RPRRAAVSSFGVSGTN 2148

Qy        421 AHVILEEAP 429
              || |:|:||
Db       2149 AHTIIEQAP 2157


RESULT 11
Q93H83
ID   Q93H83      PRELIMINARY;      PRT;  3352 AA.
AC   Q93H83;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA5.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
```

```
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69307.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002362; Antenna_beta.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00969; ANTENNA_COMP_BETA; UNKNOWN_1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3352 AA;  353395 MW;  66A5865BFD2027B7 CRC64;

  Query Match             73.1%;  Score 1656.5;  DB 2;  Length 3352;
  Best Local Similarity   73.4%;  Pred. No. 6.2e-85;
  Matches  315;  Conservative   42;  Mismatches   71;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              :|||||||||:|||| | :| | |:::  ||:| || |||||:: :|||||  |  |
Db         35 DPIAIVGMACRYPGGVRSPEDLWHLVAAGGDAVGDFPADRGWDVEGIYDPDPDARGKTYA 94

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||||:|| ||  |||||||||:|||||||||||| ||| | |||:  :: |: ||||
Db         95 RRGGFLYEAGEFDPGFFGISPREAVAMDPQQRLLLETTWETFERAGIDAESVRGSRTGVF 154

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               |:  |||   :   |  :||   || : ||:||||:|||| |||||:||||||||||||
Db        155 VGSGYQDYVNLLVGVGGESDGHLGTGNSASVMSGRIAYTFGLEGPAVTVDTACSSSLVAL 214

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              | | |||| |||||||||| ||::| |||||||||||| || |||| | |||||| |||
Db        215 HWAIQALRNGECSMALAGGVQVMTTPTAFVEFSRQRGLAPDGRCKAFGAGADGTGWAEGV 274

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||| |||| |||||||||||||||||||||||:||||||||||:||||
Db        275 GMLLVERLSDAVRNGHEVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALASAGLSPA 334

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              :| |||||||||||||||||||||||::|  : ||||||:|||:||:|||||| |:||
Db        335 QIDVVEAHGTGTTLGDPIEAQALLATYGRERTEDRPLWLGSLKSNIGHSQAAAGVGGIIK 394

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:|||:||:|||||||:||||||||:|||| | ||      ||| ||||||||||
Db        395 MVMAMRHGVLPQTLHVDEPTPHVDWSAGAVRLLTEAVEWPETH-HPRRAAVSSFGVSGTN 453

Qy        421 AHVILEEAP 429
              || |:|:||
Db        454 AHTIIEQAP 462


RESULT 12
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match             73.1%;  Score 1656;  DB 2;  Length 11096;
  Best Local Similarity   73.6%;  Pred. No. 2.9e-84;
  Matches  320;  Conservative   40;  Mismatches   67;  Indels    8;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              :|| :|||||||||||||| :| | |:    |||  |||:|||||| |||||||:| || ||
Db       3501 DPIVVVGMACRFPGGVTSPEDLWRLVDDGTDAITTFPTNRGWDLDNLYDPDPEHFGTSYT 3560

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||::|| ||  |||:|||||||| | |||||||::|| || |||:| || |: ||||
Db       3561 RSGGFLHEAGEFDPAFFGMSPREALATDSQQRLLLESSWEAIERAGIDPLTLRGSATGVF 3620

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               |    ||  :   |:  :||   |:|||| |||:||  |||||||:||||||||||||
Db       3621 AGVMYSDYGSIL--GGKEFEGFQGQGSAGSVASGRVSYALGFEGPAVTVDTACSSSLVAL 3678

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              | | |||||||||:|||||||||:|  |||||||||| ||  |||: |||| || |||
Db       3679 HWAAQALRAGECSLALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKAFAEAADGVGWSEGV 3738

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|::|| ||| |||| :|||:|||||||||||||||||||||||||||||: |||
Db       3739 GILVLERQSDAVRNGHEILAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSTA 3798

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||:||  | ||||||:|||:|||||||||||||||
Db       3799 DVDAVEAHGTGTTLGDPIEAQALLATYGRDRDPENPLWLGSLKSNIGHTQAAAGVAGVIK 3858

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:|||:||:||||| || ||||| |||:||||    ||   ||:|||||||:||||
Db       3859 MVMAMRHGVLPQTLHVDAPSSHVDWSVGAVELLTEQTAWP-ETGRVRRAGVSSFGISGTN 3917

Qy        421 AHVILE-----EAPA 430
              ||||:|     |:||
Db       3918 AHVIVEQPALVESPA 3932


RESULT 13
Q9KIV4
ID   Q9KIV4      PRELIMINARY;      PRT;  4150 AA.
AC   Q9KIV4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   8,8a-deoxyoleandolide synthase 1.
GN   OLEAI.
OS   Streptomyces antibioticus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20363406; PubMed=10908114;
RA   Shah S., Xue Q., Tang L., Carney J.R., Betlach M., McDaniel R.;
RT   "Cloning, Characterization, and Heterologous Expression of a
RT   Polyketide Synthase and P-450 Oxidase Involved in the Biosynthesis of
RT   the Antibiotic Oleandomycin.";
RL   J. Antibiot. 53:502-508(2000).
DR   EMBL; AF220951; AAF82408.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4150 AA;  435261 MW;  9383296C4C16647D CRC64;

  Query Match             72.7%;  Score 1646.5;  DB 2;  Length 4150;
  Best Local Similarity   72.8%;  Pred. No. 3e-84;
  Matches  313;  Conservative   41;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||||||: | :| ||:::: :| : | | :|||||  |||||||  || |
Db       2546 EPIAIVGMACRFPGGIASPEDLWDVLTAGEDVVSGLPQNRGWDLGRLYDPDPDRAGTSYM 2605

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | ||::|| ||| |||||||||||||||||||||||:|| :| ||| |  | |:|||||
Db       2606 REGAFLHEAGEFDAAFFGISPREALAMDPQQRLLLETSWEALERAGITPSKLAGSPTGVF 2665

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               | : |||| :  :     :|: |||:  || |||::|||| |||||:||||||||||||
Db       2666 FGMSNQDYAAQAGDVPSELEGYLLTGSISSVASGRVAYTFGLEGPAVTVDTACSSSLVAL 2725

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| | || ||||:|| |||||||||     |||||||: || ||||:|:||| |  |||
Db       2726 HLAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCKAFAASADGFGAAEGV 2785

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||||| | ||||||||||||||||||||||| |||||||||| |||:|||:
Db       2786 GVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 2845

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||||||| |    |:|||||||:||||||||||||:|
Db       2846 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRTSGRPVWLGSVKSNIGHTQAAAGVAGVMK 2905

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||:||  |::|:|||||||||||||||| |:|  | |||  | ||:|||||||||:||||
Db       2906 MVLALGRGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPWSRG-GRVRRAGVSSFGISGTN 2964

Qy        421 AHVILEEAPA 430
              ||||:|||||
Db       2965 AHVIVEEAPA 2974


RESULT 14
```

```
Q9X993
ID   Q9X993      PRELIMINARY;      PRT;  6797 AA.
AC   Q9X993;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase (PIMS1 protein).
GN   PIMS1.
OS   Streptomyces natalensis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=68242;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214571; PubMed=10187796;
RA   Aparicio J.F., Colina A.J., Ceballos E., Martin J.F.;
RT   "The biosynthetic gene cluster for the 26-membered ring polyene
RT   macrolide pimaricin. A new polyketide synthase organization encoded by
RT   two subclusters separated by functionalization genes.";
RL   J. Biol. Chem. 274:10133-10139(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Aparicio J.F., Fouces R., Mendes M.V., Olivera N.;
RT   "A complex multienzyme system encoded by five polyketide synthase
RT   genes is involved in the biosynthesis of the 26-membered polyene
RT   macrolide pimaricin in Streptomyces natalensis.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ132222; CAB41041.1; -.
DR   EMBL; AJ278573; CAC20931.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00626; RCC1_2; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6797 AA;  710149 MW;  1D2DF85074D050DA CRC64;

  Query Match             72.6%;  Score 1645.5;  DB 2;  Length 6797;
  Best Local Similarity   73.5%;  Pred. No. 6.2e-84;
  Matches  314;  Conservative   46;  Mismatches   64;  Indels    3;  Gaps    2;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              :|: ||||:||:|||| | :| | |:|   ||:  |||||||||:::|| |||:  || ||
Db       3299 DPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDPEALGTSYT 3358

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||::|  || :|||:||||||| | |||||||| || ||  ||:| :| |: ||||
Db       3359 RSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPASLRGSRTGVF 3418

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               |    ||   :   |   :|:   |:| |: |||:||||||||||:||||||||||||
Db       3419 AGVMYTDYGDLL--VGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVDTACSSSLVAL 3476

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              | | |:||:||||:|:|||||||:|  |||||||||:||| ||||: |||| ||||||
Db       3477 HWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFADAADGVGWGEGV 3536

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||::|||||| ||||||||||||||||||||||||||:||||||||||:|||||
Db       3537 GMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALASAGLSAA 3596
```

```
Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||:|  : || |||||||:|| |||:|||||||
Db       3597 DVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIK 3656

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||:|:|||:|||||||||||| |||||||||:|||    || |||  |||||||||||||
Db       3657 MVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQGEGP-RRAGVSSFGVSGTN 3715

Qy        421 AHVILEE 427
              |||||||
Db       3716 AHVILEQ 3722
```

RESULT 15
Q93NW6
ID Q93NW6 PRELIMINARY; PRT; 10917 AA.
AC Q93NW6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AmphC.
GN AMPHC.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
RT nodosus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF357202; AAK73514.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS50075; ACP_DOMAIN; 6.
DR PROSITE; PS00059; ADH_ZINC; UNKNOWN_1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
KW Phosphopantetheine.
SQ SEQUENCE 10917 AA; 1132905 MW; 15AC5956BB5810A1 CRC64;

 Query Match 72.5%; Score 1641; DB 2; Length 10917;
 Best Local Similarity 73.0%; Pred. No. 2e-83;
 Matches 316; Conservative 45; Mismatches 66; Indels 6; Gaps 3;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              :|| |||||||||| ||| :| | |:|   |||  || :||||:|:|| |||:| || ||
Db       3501 DPIVIVGMACRFPGDVTSPEDLWRLVSEGTDAISDFPANRGWDVDSLYHPDPEHSGTSYT 3560

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| || :|||:|||||||| | |||||||||:|| || |||:| :|  : ||||
Db       3561 RSGGFLHDAGDFDPDFFGMSPREALATDSQQRLLLETSWEAIERAGIDPLSLRDSRTGVF 3620

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |    ||   :   |:  :||    |:|||| |||:||||||||||||:|||||||||||
Db       3621 AGVMYHDYGSLL--GGKEFEGFQGQGSAGSVASGRVSYTFGFEGPAVTVDTACSSSLVAL 3678

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              | | ||||:||||:||||||||||:|  |:|||||:||:|||  |||: :||| || |||
Db       3679 HWAAQALRSGECSLALAGGVTVMSTPSTFIEFSRQQGLSADGRSKAFAESADGVGWSEGV 3738

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||:|| ||| |||| :|||||||||||||||||||||||||||||||||||||: ||| |
Db       3739 GMLLLERQSDAIRNGHEILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSTG 3798

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||:|||||:|  | ||||||:|||:|||||||||||
Db       3799 DVDAVEAHGTGTTLGDPIEAQALIATYGQNRTDEQPLWLGSLKSNIGHTQAAAGVAGVIK 3858

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              |||::|||:|| ||||| || ||||: |||:||||   ||  : |:||| |||||:||||
Db       3859 MVMSMRHGVLPCTLHVDAPSSHVDWTEGAVELLTEQTEWPETD-RVRRAAVSSFGISGTN 3917

Qy        421 AHVILEE---APA 430
              |||||||   |||
Db       3918 AHVILEQPVVAPA 3930
```

Search completed: June 17, 2003, 13:07:56
Job time : 20.8736 secs

THIS PAGE BL

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 4.68929 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_2086_2515
Perfect score: 2265
Sequence: 1 EPIAIVGMACRFPGGVTSAD.....VSSFGVSGTNAHVILEEAPA 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1654	73.0	3519	1	OL56_STRAT	Q07017 streptomyce
2	1544.5	68.2	3567	1	ERY2_SACER	Q03132 saccharopol
3	1495.5	66.0	3491	1	ERY1_SACER	Q03131 saccharopol
4	1425.5	62.9	3172	1	ERY3_SACER	Q03133 saccharopol
5	1104.5	48.8	1538	1	PPSB_MYCTU	Q10978 mycobacteri
6	1066	47.1	1876	1	PPSA_MYCTU	Q10977 mycobacteri
7	1051	46.4	2110	1	MCAS_MYCBO	Q02251 mycobacteri
8	865.5	38.2	4447	1	PKSK_BACSU	P40803 bacillus su
9	843	37.2	4427	1	PKSL_BACSU	Q05470 bacillus su
10	830	36.6	1774	1	MSAS_PENPA	P22367 penicillium
11	793	35.0	4273	1	PKSM_BACSU	P40872 bacillus su
12	713	31.5	2511	1	FAS_CHICK	P12276 gallus gall
13	702.5	31.0	2181	1	STCA_EMENI	Q12397 emericella
14	684.5	30.2	2109	1	PKS1_ASPPA	Q12053 aspergillus
15	682	30.1	2505	1	FAS_RAT	P12785 rattus norv
16	681	30.1	1986	1	WA_EMENI	Q03149 emericella
17	674.5	29.8	2504	1	FAS_HUMAN	P49327 homo sapien
18	628	27.7	352	1	FAS_ANSAN	P36189 anser anser
19	370.5	16.4	401	1	NOE4_RHIME	P06231 rhizobium m
20	368	16.2	402	1	NODE_RHIME	P06230 rhizobium m
21	365	16.1	426	1	KAS1_STRGA	P16538 streptomyce
22	354	15.6	402	1	NODE_RHIS3	P72331 rhizobium s
23	352	15.5	403	1	NODE_RHILV	P04683 rhizobium l
24	343.5	15.2	424	1	KASA_STRCO	Q02059 streptomyce
25	342	15.1	421	1	FABF_RHIME	P56902 rhizobium m
26	337.5	14.9	401	1	NODE_RHILT	P04684 rhizobium l
27	332	14.7	412	1	FABF_ECOLI	P39435 escherichia
28	325	14.3	420	1	KAS1_STRCM	P41175 streptomyce
29	319	14.1	417	1	FAB2_MYCTU	Q10525 mycobacteri
30	317.5	14.0	422	1	KAS1_STRCN	Q02578 streptomyce
31	317.5	14.0	442	1	CEM1_YEAST	P39525 saccharomyc
32	312.5	13.8	422	1	KAS1_STRHA	Q05356 streptomyce
33	308.5	13.6	413	1	FABF_VIBCH	Q9kqh9 vibrio chol
34	307	13.6	414	1	FABF_VIBHA	P55338 vibrio harv
35	304	13.4	421	1	KAS1_STRVN	P16540 streptomyce
36	292.5	12.9	462	1	FABB_HORVU	P23902 hordeum vul
37	290.5	12.8	420	1	FAB2_MYCLE	O69473 mycobacteri
38	288	12.7	416	1	FAB1_MYCLE	Q9cbs7 mycobacteri
39	286	12.6	416	1	FAB1_MYCTU	Q10524 mycobacteri
40	281.5	12.4	406	1	FABB_ECOLI	P14926 escherichia
41	280.5	12.4	416	1	FABF_SYNY3	P73283 synechocyst
42	279	12.3	473	1	FABB_ARATH	P52410 arabidopsis
43	277.5	12.3	423	1	KAS1_STRCO	P23155 streptomyce
44	261.5	11.5	406	1	FABB_HAEIN	P43710 haemophilus
45	256	11.3	416	1	PKSF_BACSU	P40804 bacillus su

ALIGNMENTS

RESULT 1
OL56_STRAT
ID OL56_STRAT STANDARD; PRT; 3519 AA.
AC Q07017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oleandomycin polyketide synthase, modules 5 and 6.
GN ORFB.
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150470; PubMed=8107683;
RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT "Characterisation of a Streptomyces antibioticus gene encoding a type
RT I polyketide synthase which has an unusual coding sequence.";
RL Mol. Gen. Genet. 242:358-362(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC LACTONE RING.
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; L09654; AAA19695.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF00550; pp-binding; 2.
DR Pfam; PF00698; Acyl_transf; 2.
DR Pfam; PF00975; Thioesterase; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT DOMAIN 1 ? MODULE 5.
FT DOMAIN ? 3519 MODULE 6.
FT DOMAIN 32 501 BETA-KETOACYL SYNTHASE 1.
FT DOMAIN 569 890 ACYLTRANSFERASE (AT) 1.
FT DOMAIN 1200 1382 BETA-KETOACYL REDUCTASE 1.
FT DOMAIN 1487 1561 ACYL CARRIER (ACP) 1.
FT DOMAIN 1686 2156 BETA-KETOACYL SYNTHASE 2.

```
FT   DOMAIN      2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN      2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN      3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE       ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING        ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN      3270   3519       THIOESTERASE.
FT   ACT_SITE     210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE     660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     1203   1249       NADP.
FT   BINDING     1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     2859   2905       NADP.
FT   BINDING     3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             73.0%;  Score 1654;  DB 1;  Length 3519;
  Best Local Similarity   73.3%;  Pred. No. 8.8e-106;
  Matches  315;  Conservative   40;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              | |||  |:||| ||: | :| |  ::  :||: | | ||||||| || |||::||| |
Db       1687 EAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGTTYV 1746

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | | |||  ||| |||||||||||||||||||||:||  | | |:|:|: || ||:|
Db       1747 REGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTATGIF 1806

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |   | |       | :|  |: ||||| :|:|||||||| |||||:||||||||||||
Db       1807 IGAGHQGYGPDPKRAPESVAGYLLTGTASAVLSGRISYTFGLEGPAVTVDTACSSSLVAL 1866

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||||:|:|||| |||:| |||||||:|:| || ||||:||||| ||||||
Db       1867 HLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAADGMGWGEGV 1926

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              :||:|||||| | |||||||||||||||||||| |||||||||| |||:|||:
Db       1927 SLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 1986

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| |||||||||||||||| |||   |:||||||:|||||||||||:|
Db       1987 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAAAGVAGVMK 2046

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||:||  |:::|:|||||||||||||||||:|||| ||    |||||:|:|||||||
Db       2047 MVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGISAFGVSGTN 2106

Qy        421 AHVILEEAPA 430
              ||||:|||||
Db       2107 AHVIVEEAPA 2116


RESULT 2
ERY2_SACER
ID   ERY2_SACER     STANDARD;      PRT;  3567 AA.
AC   Q03132; Q54096;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE   deoxyerythronolide B synthase II) (DEBS 2).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN         1   1484       MODULE 3.
FT   DOMAIN      1485   3567       MODULE 4.
FT   DOMAIN        27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN       559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                 NON-FUNCTIONAL).
FT   DOMAIN      1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN      1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN      2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN      2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN      3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN      3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE     202    202       THIOESTER BOND.
FT   ACT_SITE     651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    1661   1661       THIOESTER BOND.
FT   ACT_SITE    2115   2115       ACYL-ENZYME INTERMEDIATE.
```

```
FT   NP_BIND     2961   2978       NADP (ER).
FT   NP_BIND     3142   3157       NADP (KR).
FT   BINDING     3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT     438    438       R -> A (IN REF. 2).
FT   CONFLICT     480    480       T -> S (IN REF. 2).
FT   CONFLICT    1241   1241       L -> F (IN REF. 2).
FT   CONFLICT    2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             68.2%;  Score 1544.5;  DB 1;  Length 3567;
  Best Local Similarity   68.8%;  Pred. No. 3.1e-98;
  Matches  296;  Conservative   47;  Mismatches   84;  Indels    3;  Gaps    2;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||: ||||||: | :  | :::   :   ||| |||||:  || ||||:||| |
Db       1489 EPIAIVGIGCRFPGGIGSPEQLWRVLAEGANLTTGPPADRGWDIGRLYHPDPDNPGTSYV 1548

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |||| ||  ||  ||||:|||||||||||||:|||||| :| |||:|  | || ||||
Db       1549 DKGGFLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGIDPDALRGTDTGVF 1608

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              | ||| |   :   :  ||:   | : ||:||||:||||:||||::|||||||||| :
Db       1609 VGMNGQSYMQLLAGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPALTVDTACSSSLVGI 1668

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||||:|||||||||| |  ||:|| |||||:|| |||||| |||    |||
Db       1669 HLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKAFSARADGFALSEGV 1728

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |::| || |  |||:||||:|||||||||||||| |||||||:|||||||| :|: |
Db       1729 AALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQERVIRQALAASGVPAA 1788

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||| ||:|||||||  : || |||||:|:|||||||| |||||
Db       1789 DVDVVEAHGTGTELGDPIEAGALIATYGQDR--DRPLRLGSVKTNIGHTQAAAGAAGVIK 1846

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              :|:|:|||:|||:|| || |||:|| :|||::| | |||| || | ||||||||||||||
Db       1847 VVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGE-RPRRAGVSSFGVSGTN 1905

Qy        421 AHVILEEAPA 430
              ||||:|||||
Db       1906 AHVIVEEAPA 1915


RESULT 3
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN         1   1972       MODULE 1.
FT   DOMAIN      1979   3491       MODULE 2.
FT   DOMAIN         1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN       414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN       503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN      1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN      1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN      1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN      2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN      3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN      3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE     145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING      447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE     677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     1614   1660       NADP.
FT   BINDING     1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     3058   3104       NADP.
FT   BINDING     3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             66.0%;  Score 1495.5;  DB 1;  Length 3491;
```

```
  Best Local Similarity   66.2%;  Pred. No. 7e-95;
  Matches  284;  Conservative   54;  Mismatches   88;  Indels    3;  Gaps    2;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|:| |||| ||||:: ::||:|:|  :||: | |||||||||:|: |||   || :
Db        505 EPVAVVAMACRLPGGVSTPEEFWELLSEGRDAVAGLPTDRGWDLDSLFHPDPTRSGTAHQ 564

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|  ||  |||:|||||||:||||||:|| :|| :| ||| | :|  :|||||
Db        565 RGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLERAGIPPTSLQASPTGVF 624

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |   |:|  |:   |:  :|: :|||  || ||||:|| | ||||:|||||||||||:
Db        625 VGLIPQEYGPRLAEGGEGVEGYLMTGTTSVASGRIAYTLGLEGPAISVDTACSSSLVAV 684

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:|| || |:|:|||||||| :||  |:|||   || || |||||| |:| |  ||
Db        685 HLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAFSAGANGFGMAEGA 744

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ||||:|||||| |||| ||||:||:||| |||||||:|||| :| |||:|||| :||
Db        745 GMLLLERLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRVIQQALAESGLGPA 804

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |:|||||||||| |||||||:||   ||:||  | || |||||||:|||||||||||||
Db        805 DIDAVEAHGTGTRLGDPIEAARALFEAYGRDR--EQPLHLGSVKSNLGHTQAAAGVAGVIK 862

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||:|:| | |||||  | |  :|||:||: || |  ||| |  | |||||||||:||||
Db        863 MVLAMRAGTLPRTLHASERSKEIDWSSGAISLLDEPEPWPAG-ARPRRAGVSSFGISGTN 921

Qy        421 AHVILEEAP 429
              || |:||||
Db        922 AHAIIEEAP 930


RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
```

```
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match          62.9%;  Score 1425.5;  DB 1;  Length 3172;
  Best Local Similarity   65.3%;  Pred. No. 4.1e-90;
  Matches  280;  Conservative   44;  Mismatches   92;  Indels   13;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||||| | | : ||: :|   |||   | ||||      :||||
Db         39 EPIAIVGMACRFPGDVDSPESFWEFVSGGGDAIAEAPADRGW------EPDPD------A 86

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | || |  || ||| |||||||||||||||||::|| :|| :| || :| :| |: ||||
Db         87 RLGGMLAAAGDFDAGFFGISPREALAMDPQQRIMLEISWEALERAGHDPVSLRGSATGVF 146

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    ||  |   |      |:  |||| || |||::|  | ||||::|||||| | ||
Db        147 TGVGTVDYGPRPDEAPDEVLGYVGTGTASSVASGRVAYCLGLEGPAMTVDTACSSGLTAL 206

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| ::||  || :|||||||||||||| ||  | |||||| || || |||| |  ||
Db        207 HLAMESLRRDECGLALAGGVTVMSSPGAFTEFRSQGGLAADGRCKPFSKAADGFGLAEGA 266

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|:::||| | | |  |||| ||||||||||||||:||:||||||:|| |||: ||
Db        267 GVLVLQRLSAARREGRPVLAVLAGSAVNQDGASNGLTAPSGPAQQRVIRRALENAGVRAG 326

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||  |||:||| :|  : |||:||||||:||||||||||||:|
Db        327 DVDYVEAHGTGTRLGDPIEVHALLSTYGAERDPDDPLWIGSVKSNIGHTQAAAGVAGVMK 386

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              |:||||| :||||| ||||| ::|  ||| ::::   || || | |||||||||:||||
Db        387 AVLALRHGEMPRTLHFDEPSPQIEWDLGAVSVVSQARSWPAGE-RPRRAGVSSFGISGTN 445

Qy        421 AHVILEEAP 429
              |||||:||||
Db        446 AHVIVEEAP 454


RESULT 5
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
```

```
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             48.8%;  Score 1104.5;  DB 1;  Length 1538;
  Best Local Similarity   52.5%;  Pred. No. 2.2e-68;
  Matches  224;  Conservative   64;  Mismatches  134;  Indels    5;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|:||: ||||| |    : ||| : : ::||   | || || :  | |||  ||   |
Db         34 EPVAVVGIGCRFPGDVDGPESFWDFLVAGRNAISTVPADR-WDAEAFYHPDPLTPGRMTT 92

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              : |||: |   ||||||||:|||| |||||||:||| ||| :||||| | :| || | |
Db         93 KWGGFVPDVAGFDAEFFGITPREAAAMDPQQRMLLEVAWEALEHAGIPPDSLGGTRTAVM 152

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |    :|   :  : |: | :: || | |:  |||||  |  ||||:||||||||||:
Db        153 MGVYFNEYQSMLAASPQNVDAYSGTGNAHSITVGRISYLLGLRGPAVAVDTACSSSLVAV 212

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:||  |  :|||||::   |   :  |    |:  | | || |||||   |||
Db        213 HLACQSLRLRETDLALAGGVSITLRPETQIAISAWGLLSPQGRCAAFDAAADGFVRGEGA 272

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:::::||:|| |:|  :|||||||||||||| |||:|||| :|  ||  || :  ::
Db        273 GVVVLKRLTDAVRDGDQVLAVVRGSAVNQDGRSNGVTAPNTAAQCDVIADALRSGDVAPD 332

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |: |||||||| |||||| :|| ||||    |     ||:||:|:|| :||||:|| ||
Db        333 SVNYVEAHGTGTVLGDPIEFEALAATYGH---GGDACALGAVKTNIGHLEAAAGIAGFIK 389

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
                :|::   :|  ||  : :| :| ::    : |:  |||  ||  ||| |||||: |||
Db        390 ATLAVQRATIPPNLHFSQWNPAIDAASTRFFVPTQNSPWPTAEGP-RRAAVSSFGLGGTN 448

Qy        421 AHVILEE 427
              ||||:|:
Db        449 AHVIIEQ 455


RESULT 6
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             47.1%;  Score 1066;  DB 1;  Length 1876;
  Best Local Similarity   50.0%;  Pred. No. 1.2e-65;
  Matches  215;  Conservative   66;  Mismatches  141;  Indels    8;  Gaps    4;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCY- 59
              ||||:||| ||||||::  :  || :  : :|   |  | |     ::  |
Db        102 EPIAVVGMGCRFPGGISCPEALWDFLCERRSSISQVPPQR-WQ---PFEGGPPEVAAALA 157

Qy         60 --TRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPT 117
                || | || |   |||||| ||| ||  |||||||||| ||| :||||| | ||  : |
Db        158 RTTRWGSFLPDIDAFDAEFFEISPSEADKMDPQQRLLLEVAWEALEHAGIPPGTLRRSAT 217

Qy        118 GVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSL 177
              |||  |    :|           ||:: :| | |:|: |:||     ||:|:|||||||||
Db        218 GVFAGACLSEYGAMASADLSQVDGWSNSGGAMSIIANRLSYFLDLRGPSVAVDTACSSSL 277

Qy        178 VALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWG 237
              ||:|||||:||  :| :|:| || :: ||  |  | :   |:  | |:|| | |||   |
Db        278 VAIHLACQSLRTQDCHLAIAAGVNLLLSPAVFRGFDQVGALSPTGQCRAFDATADGFVRG 337

Qy        238 EGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGL 297
              || |:::::||:|| |:| |||||: |||||||| |||| |||  :|  |:| |  |||:
Db        338 EGAGVVVLKRLTDAQRDGDRVLAVICGSAVNQDGRSNGLMAPNPAAQMAVLRAAYTNAGM 397
```

```
Qy        298 SAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAG 357
              :|| |||||||| |||||||:||    |: |  : || :||||:|:|||:||||:||
Db        398 QPSEVDYVEAHGTGTLLGDPIEARALGTVLGRGRPEDSPLLIGSVKTNLGHTEAAAGIAG 457

Qy        358 VIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVS 417
              || |:|::|| :|   | :  :||: ::   ::::    ||   |  |||||||||
Db        458 FIKTVLAVQHGQIPPNQHFETANPHIPFTDLRMKVVDTQTEWP-ATGHPRRAGVSSFGFG 516

Qy        418 GTNAHVILEE 427
              ||||||::|:
Db        517 GTNAHVVIEQ 526
```

RESULT 7
MCAS_MYCBO
ID MCAS_MYCBO STANDARD; PRT; 2110 AA.
AC Q02251;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Mycocerosic acid synthase.
GN MAS.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92406887; PubMed=1527058;
RA Mathur M., Kolattukudy P.E.;
RT "Molecular cloning and sequencing of the gene for mycocerosic acid
RT synthase, a novel fatty acid elongating multifunctional enzyme, from
RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL J. Biol. Chem. 267:19388-19395(1992).
CC -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC FORM MYCOCEROSYL LIPIDS.
CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC ARRANGEMENT.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M95808; AAA25369.1; -.
DR PIR; B44110; B44110.
DR HSSP; P73283; 1E5M.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF00550; pp-binding; 1.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT DOMAIN 1 430 BETA-KETOACYL SYNTHASE.
FT DOMAIN 533 852 ACYL TRANSFERASE.
FT DOMAIN ? ? ENOYL REDUCTASE.
FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.
FT DOMAIN 2026 2096 ACYL CARRIER (ACP).
FT ACT_SITE 177 177 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 623 623 ACYL TRANSFERASES (BY SIMILARITY).
FT NP_BIND 1561 1578 NADP (ER).
FT NP_BIND 1765 1780 NADP (KR).
FT BINDING 2059 2059 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 2110 AA; 225577 MW; CC658215D7155300 CRC64;

 Query Match 46.4%; Score 1051; DB 1; Length 2110;
 Best Local Similarity 50.3%; Pred. No. 1.5e-64;
 Matches 218; Conservative 61; Mismatches 142; Indels 12; Gaps 7;

```
Qy          2 PIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTR 61
              |:|::|| || |||: | |  |: :   | :   | || || |  |||:|  ||   :|
Db          7 PVAVIGMGCRLPGGINSPDKLWESLLRGDDLVTEIPPDR-WDADDYYDPEPGVPGRSVSR 65

Qy         62 NGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFT 121
              |||| |   |||||||| ||| ::||||||||||:|| |||||::| :| |: | |||
Db         66 WGGFLDDVAGFDAEFFGISEREATSIDPQQRLLLETSWEAIEHAGLDPASLAGSSTAVFT 125

Qy        122 GTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
              |   :|| :    ||     : :||   || ||||::| |  |||:: |||||  |:|:|
Db        126 GLTHEDYLVLTTTAGGLASPYVVTGLNNSVASGRIAHTLGLHGPAMTFDTACSSGLMAVH 185

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              |||::|  ||  :|||||  |:  | | |  | |  |:: | | :| | |||    ||
Db        186 LACRSLHDGERDLALAGGCAVLLEPHACVAASAQGMLSSTGRCHSFDADADGFVRSEGCA 245

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
              |:|::|| || |:|:|: |||||:| |||| :  || |:  :|  | | ||| ||:
Db        246 MVLLKRLPDALRDGNRIFAVVRGTATNQDGRTETLTMPSEDAQVAVYRAALAAAGVQPET 305

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              |  |||||||| :|||| ::|   ||   ||  |  ||| |||:||: |:||  |:||
Db        306 VGVVEAHGTGTPIGDPIEYRSLARVYG---AGT-PCALGSAKSNMGHSTASAGTVGLIKA 361

Qy        362 VMALRHGLLPRTLHVDE-PSPHVDWSAG--AVQLLTETVPWPGGEGRL-RRAGVSSFGVS 417
              :::||||::|  || :  |   |  |   | :|  |||  |     :|  |||||:|
Db        362 ILSLRHGVVPPLLHFNRLPDELSDVETGLFVPQAVT---PWPNGNDHTPKRVAVSSFGMS 418

Qy        418 GTNAHVILEEAPA 430
              ||| | |:|||||
Db        419 GTNVHAIVEEAPA 431
```

RESULT 8
PKSK_BACSU
ID PKSK_BACSU STANDARD; PRT; 4447 AA.
AC P40803;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksK (PKS).
GN PKSK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PB1424;
RX MEDLINE=95219083; PubMed=7704258;
RA Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;
RT "Sequence around the 159 degree region of the Bacillus subtilis
RT genome: the pksX locus spans 33.6 kb.";
RL Microbiology 141:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

```
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11039; AAA85144.1; -.
DR   EMBL; Z99112; CAB13590.1; -.
DR   EMBL; Z99113; CAB13601.1; -.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10930; pksK.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF00668; Condensation; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;
KW   Complete proteome.
FT   DOMAIN       1     68       ACYL CARRIER (ACP) 1.
FT   DOMAIN    1063   1130       ACYL CARRIER (ACP) 2.
FT   DOMAIN    2516   2589       ACYL CARRIER (ACP) 3.
FT   DOMAIN    2618   2687       ACYL CARRIER (ACP) 4.
FT   DOMAIN    3868   3937       ACYL CARRIER (ACP) 5.
FT   BINDING     31     31       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   1093   1093       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   2552   2552       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   2650   2650       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE  2915   2915       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING   3900   3900       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE  4147   4147       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
SQ   SEQUENCE   4447 AA;  496058 MW;  9C6DB4A6C0A9C057 CRC64;

  Query Match             38.2%;  Score 865.5;  DB 1;  Length 4447;
  Best Local Similarity   41.8%;  Pred. No. 2.1e-51;
  Matches  185;  Conservative   80;  Mismatches  159;  Indels   19;  Gaps    8;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLY-DPDPDHPGTCY 59
              ||:||||:: |||| :   |:||  :   :|:|   | || ||    | :|| |   |
Db       2744 EPVAIVGISGRFPGAM-DIDEFWKNLEEGKDSITEVPKDR-WDWREHYGNPDTDVNKT-D 2800

Qy         60 TRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGV 119
               : |||:     ||  |||||||||  :|||||||:   |: :| || :| :| || ||:
Db       2801 IKWGGFIDGVAEFDPLFFGISPREADYVDPQQRLLMTYVWKALEDAGCSPQSLSGTGTGI 2860

Qy        120 FTGTNGQDYALRVHNAGQSTDGFALTG-TAGSVISGRISYTFGFEGPAVSVDTACSSSLV 178
              | ||    |   | |    :| | ||    ||   |:||     ||:  |:||||||||
Db       2861 FIGTGNTGYKDLFHRANLPIEGHAATGHMIPSVGPNRMSYFLNIHGPSEPVETACSSSLV 2920

Qy        179 ALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGE 238
              |:| |  |:: |:| ||:|||| : :  | : :|:   |: || || ||| |:|   ||
Db       2921 AIHRAVTAMQNGDCEMAIAGGVNTILTEEAHISYSKAGMLSTDGRCKTFSADANGYVRGE 2980

Qy        239 GVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLS 298
              ||||:::::| || |:|: :  |:||:| |  | :| ||:||  :|  :: :|   | :
Db       2981 GVGMVMLKKLEDAERDGNHIYGVIRGTAENHGGRANTLTSPNPKAQADLLVRAYRQADID 3040

Qy        299 AGDVDAVEAHGTGTTLGDPIEAQALLATYGQ--DRAGEGP------LWLGSVKSNVGHT 349
                |  :|||||| ||||||   | | : :  :  ||           :||||||:||
Db       3041 PSTVTYIEAHGTGTELGDPIEINGLKAAFKELSNMRGESQPDVPDHRCGIGSVKSNIGHL 3100

Qy        350 QAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPG-----GEG 404
              : |||::|:||::: ::|  | ::|| :  :|::  :    .:: |   |       |
Db       3101 ELAAGISGLIKVLLQMKHKTLVKSLHCETLNPYLQLTDSPFYIVQEKQEWKSVTDRDGNE 3160

Qy        405 RLRRAGVSSFGVSGTNAHVILEE 427
                ||||:||||: | |||:::||
Db       3161 LPRRAGISSFGIGGVNAHIVIEE 3183


RESULT 9
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
```

```
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;

  Query Match             37.2%;  Score 843;  DB 1;  Length 4427;
  Best Local Similarity   41.0%;  Pred. No. 7.2e-50;
  Matches  175;  Conservative   75;  Mismatches  171;  Indels    6;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              : |||||: |:| |  :  ::|| :   ::||   || | ||:|  |||  :  |  |
Db       2763 DAIAIVGMSGRYP-GARNVREYWDNLVHARNAIRDIPTSR-WDVDKYYDPVLNKKGKVYC 2820

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              :: | | |  |||  || | | ||  |||| |: |:  ::  | || |  ||:    ||:
Db       2821 KSMGMLDDIEHFDPLFFNIPPSEAELMDPQHRIFLQEGYKAFEDAGYNARTLNEKKCGVY 2880

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               |   :| : ::   ::       || : :: : || |    :|||: :|||||||||
Db       2881 LGIMSNEYGVMLNRQSRAN----ATGNSFAIAAARIPYFLNLKGPAIPIDTACSSSLVGT 2936

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||   |  ||| |||::  :| ::: |: || ||||   |:|   |||
Db       2937 HLARQALINKEIDMALVGGVSLYLTPESYMSMCEAGMLSPDGQCKAFDNGANGFVPGEGA 2996

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              | |:::|| ||  :   :  :: || :|||| :||:|||:  ||  : |       |:
Db       2997 GALVLKRLKDAEADRDHIYGIIIGSGINQDGKTNGITAPSAKSQMDLERDIYETYGIHPE 3056

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              :  || |||||  ||||| :||   : :    :   :|||||:||| |||||||| |
Db       3057 SISYVEMHGTGTKQGDPIELEALSTVFQEKTDKKQFCAIGSVKSNIGHTSAAAGVAGVQK 3116

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ::: : |  |  ||:   |: | ::    : : ||  ||   :|: ||| ||||| ||||
Db       3117 VLLCMNHKTLVPTLNFTTPNEHFEFEHSPLYVNTELKPWETADGKPRRACVSSFGYSGTN 3176

Qy        421 AHVILEE 427
              ||:::||
Db       3177 AHIVIEE 3183


RESULT 10
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
```

```
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             36.6%;  Score 830;  DB 1;  Length 1774;
  Best Local Similarity   39.2%;  Pred. No. 2e-49;
  Matches  175;  Conservative   70;  Mismatches  157;  Indels   44;  Gaps    3;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFP----------------------TDRG 41
              :|:||||||  ||  : :  |  : |:: |:|  |                        | ||
Db         35 VAVVGMACRVAGGNHNPELLWQSLLSQKSAMGEIPPMRWEPYYRRDARNEKFLKNTTSRG 94

Qy         42 WDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWET 101
              : || | |                        || :||||||:||  ||||||: || | |
Db         95 YFLDRLED----------------------FDCQFFGISPKEAEQMDPQQRVSLEVASEA 132

Qy        102 IEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFG 161
              :| |||   :| |: | || | |  ||:  |     : : :   |||   :  ||||
Db        133 LEDAGIPAKSLSGSDTAVFWGVNSDDYSKLVLEDLPNVEAWMGIGTAYCGVPNRISYHLN 192

Qy        162 FEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD 221
                ||: :|| ||:|||||:|   ||:| ||  :|: |||  :  ||      :  :::|
Db        193 LMGPSTAVDAACASSLVAIHHGVQAIRLGESKVAIVGGVNALCGPGLTRVLDKAGAISSD 252

Qy        222 GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNG 281
              | ||:|   | |   ||| | |::: |  |  :   ||||::|||| ||| :||: |||
Db        253 GSCKSFDDDAHGYARGEGAGALVLKSLHRALLDHDNVLAVIKGSAVCQDGKTNGIMAPNS 312

Qy        282 PSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGS 341
              :||     ||: | :    |  |||| | | |||| |  |: : || ||  : | ::||
Db        313 VAQQLAANNALSAANIDPHTVRYVEAHATSTPLGDPTEISAIASVYGADRPADDPCYIGS 372

Qy        342 VKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPG 401
              :| |:|| :| ||| | || |:|:: |:||   :: : :  :||    |::: |  |||
Db        373 IKPNIGHLEAGAGVMGFIKAVLAIQKGVLPPQANLTKLNSRIDWKTAGVKVVQEATPWPE 432

Qy        402 GEGRLRRAGVSSFGVSGTNAHVILEE 427
                :  :||||| |:|  || :| ::||
```

```
Db        433 SD-PIRRAGVCSYGYGGTVSHAVIEE 457


RESULT 11
PKSM_BACSU
ID   PKSM_BACSU     STANDARD;      PRT;  4273 AA.
AC   P40872; O31781;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksM.
GN   PKSM OR PKSY.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE OF 1-1763 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Tognoni A., Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99113; CAB13603.1; -.
DR   EMBL; Z35133; CAA84505.1; -.
DR   SubtiList; BG10931; pksM.
```

```
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001601; Methyltransf.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      295    364       ACYL CARRIER (ACP) 1.
FT   DOMAIN      396    834       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     2190   2258       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2322   2737       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     3532   3947       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3410   3483       ACYL CARRIER (ACP) 3.
FT   DOMAIN     4140   4209       ACYL CARRIER (ACP) 4.
FT   BINDING     327    327       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2222   2222       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE   2476   2476       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING    3446   3446       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE   3690   3690       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING    4172   4172       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    103    103       E -> V (IN REF. 2).
FT   CONFLICT    276    276       Q -> E (IN REF. 2).
FT   CONFLICT    289    289       T -> S (IN REF. 2).
SQ   SEQUENCE   4273 AA;  477459 MW;  3BBFCF1A250AEB5A CRC64;

  Query Match          35.0%;  Score 793;  DB 1;  Length 4273;
  Best Local Similarity   39.1%;  Pred. No. 1.9e-46;
  Matches  168;  Conservative   82;  Mismatches  158;  Indels   22;  Gaps    5;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :|::|:: |:|    | |:|||: :   :: |    | || ||     || :    |: ||:
Db       2322 VAVIGISGRYPQAET-AEDFWNNLKEGRNCIEEIPKDR-WDWKAYYDKEKGKEGSIYTKW 2379

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL-HGTPTGVFT 121
              |||: |   ||  || ||| ||  ||||:|| |:||: :|| ||  | :|       |||
Db       2380 GGFIKDMDKFDPLFFQISPLEAERMDPQERLFLQTAYASIEDAGYTPDSLCSSRKIGVFA 2439

Qy        122 GTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
              |   ::|                ||     |: ||||    |:||:::|||||||| |:|
Db       2440 GVMNKNYP---------------TGYGYWSIANRISYLLNFQGPSLAVDTACSSSLTAIH 2484

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              || ::: :|    |:|||| ::  |  :   |    |:|   ||:|   |||   |||||
Db       2485 LALESIYSGSSDCAIAGGVNLVVDPVHYQNLSVMNMLSASDTCKSFGDDADGFVDGEGVG 2544

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
               :::: |  |  :|  :  |:: ||:|  | :|| | ||  :| :||::|:  | : |
Db       2545 AIVLKPLQQAIADGDHIYGVIKASAINSGGKTNGYTVPNPHAQAQVIKEAIERADIPART 2604

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              :  :|||||| ||||||   |   : :|   :    :|| |||:|| ::|||:||: |:
Db       2605 ISYLEAHGTGTALGDPIEIAGLTKAFEKDTQEKQFCAIGSSKSNIGHCESAAGIAGLTKI 2664

Qy        362 VMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP----GGEGRLRRAGVSSFGVS 417
              :   ::| :  :||    :|::::|     :  :   |     ||:   ||||:||||
Db       2665 LFQFKYGQIAPSLHAQRLNPNIEFSHTPFVVQQQLGEWKRPVIGGQEVPRRAGLSSFGAG 2724

Qy        418 GTNAHVILEE 427
              |:|||:|||| 
Db       2725 GSNAHIILEE 2734


RESULT 12
FAS_CHICK
ID   FAS_CHICK      STANDARD;      PRT;  2511 AA.
AC   P12276;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=White leghorn; TISSUE=Liver;
RX   MEDLINE=95031085; PubMed=7944406;
RA   Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Amino-terminal blocking group and sequence of the animal fatty acid
RT   synthase.";
RL   Arch. Biochem. Biophys. 314:45-49(1994).
RN   [2]
RP   SEQUENCE OF 75-1775 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89282777; PubMed=2734291;
RA   Holzer K.P., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of chicken liver fatty acid
RT   synthase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN   [3]
RP   SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89139426; PubMed=2917973;
RA   Chirala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,
RA   Wakil S.J.;
RT   "A novel cDNA extension procedure. Isolation of chicken fatty acid
RT   synthase cDNA clones.";
RL   J. Biol. Chem. 264:3750-3757(1989).
RN   [4]
RP   SEQUENCE OF 1752-2512 FROM N.A.
RX   MEDLINE=88320436; PubMed=2842766;
RA   Yuan Z., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of DNA complementary to chicken
RT   liver fatty acid synthase mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN   [5]
RP   SEQUENCE OF 2202-2512 FROM N.A.
RX   MEDLINE=89088152; PubMed=3207710;
RA   Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT   "Characterization of a genomic and cDNA clone coding for the
RT   thioesterase domain and 3' noncoding region of the chicken liver
RT   fatty acid synthase gene.";
RL   Biochemistry 27:7778-7785(1988).
RN   [6]
RP   SEQUENCE OF 2121-2209.
RX   MEDLINE=89192401; PubMed=2648999;
RA   Huang W.-Y., Stoops J.K., Wakil S.J.;
RT   "Complete amino acid sequence of chicken liver acyl carrier protein
RT   derived from the fatty acid synthase.";
RL   Arch. Biochem. Biophys. 270:92-98(1989).
RN   [7]
RP   SEQUENCE OF 2209-2508.
RC   STRAIN=White leghorn;
RX   MEDLINE=89088151; PubMed=3207709;
RA   Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Complete amino acid sequence of the thioesterase domain of chicken
RT   liver fatty acid synthase.";
RL   Biochemistry 27:7773-7777(1988).
RN   [8]
RP   SEQUENCE OF 667-674 AND 1698-1709.
RX   MEDLINE=89323081; PubMed=2751995;
RA   Chang S.I., Hammes G.G.;
RT   "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
RT   fluorescence resonance energy transfer in chicken liver fatty acid
RT   synthase.";
RL   Biochemistry 28:3781-3788(1989).
```

```
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
CC       POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04485; AAB46389.1; -.
DR   EMBL; J03860; AAA48767.1; -.
DR   EMBL; J02839; AAA82106.1; ALT_SEQ.
DR   PIR; A33918; XYCHFA.
DR   PIR; A32015; A32015.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;
KW   Alternative splicing; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   DOMAIN        1   ?411       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      427    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1638   1866       ENOYL REDUCTASE.
FT   DOMAIN     1867   2119       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2124   2180       ACYL CARRIER (ACP).
FT   DOMAIN     2209   2511       THIOESTERASE.
FT   MOD_RES       1      1       ACETYLATION.
FT   ACT_SITE    160    160       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    579    579       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   ACT_SITE    877    877       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   NP_BIND    1674   1691       NADP (ER).
FT   BINDING    1707   1707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1888   1903       NADP (KR).
FT   BINDING    2157   2157       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2308   2308       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2481   2481       THIOESTERASE (BY SIMILARITY).
FT   VARSPLIC   2348   2348       T -> TQCFSFSLF (IN ISOFORM 1).
FT   CONFLICT     77     78       QL -> PV (IN REF. 2).
FT   CONFLICT    116    116       L -> A (IN REF. 2).
FT   CONFLICT    675    675       R -> S (IN REF. 2).
FT   CONFLICT   1169   1169       K -> N (IN REF. 2).
FT   CONFLICT   1178   1178       A -> T (IN REF. 2).
FT   CONFLICT   1191   1191       R -> H (IN REF. 2).
FT   CONFLICT   1198   1198       P -> L (IN REF. 2).
FT   CONFLICT   1286   1287       DN -> ND (IN REF. 2).
FT   CONFLICT   1372   1372       K -> E (IN REF. 2).
FT   CONFLICT   1533   1533       C -> Y (IN REF. 2).
FT   CONFLICT   1577   1577       W -> R (IN REF. 2).
FT   CONFLICT   1685   1696       QAAIAIALSMGC -> ASSHCHRLEHGLA
FT                                (IN REF. 2; AAA48767).
FT   CONFLICT   1732   1732       Q -> E (IN REF. 2).
FT   CONFLICT   1745   1745       S -> N (IN REF. 2).
SQ   SEQUENCE   2511 AA;  274648 MW;  622039DAC8315D3F CRC64;

  Query Match             31.5%;  Score 713;  DB 1;  Length 2511;
  Best Local Similarity   36.9%;  Pred. No. 3.3e-41;
  Matches  158;  Conservative   83;  Mismatches  159;  Indels   28;  Gaps    9;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTC-Y 59
              | : | |:| : |    :  :||: : :  | :    || |           ||
Db          1 EDVVIAGIAGKLPES-ENLQEFWENLLNGVDMV--TEDDRRW-----------KPGIYGL 46

Qy         60 TRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGV 119
               :  | | |   ||| ||||: |::|  |||| ||||| ::| |   ||||  | || |||
Db         47 PKRNGKLKDIKKFDASFFGVHPKQAHTMDPQLRLLLEVSYEAILDGGINPTALRGTDTGV 106

Qy        120 FTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179
              : | :| :    :    :   |:::||   :::: |||| : | ||:::||||||||:|
Db        107 WVGASGSEALEALSQDPEELLGYSMTGCQRAMLANRISYFYDFTGPSLTIDTACSSSLMA 166

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239
              |  | :|:| |:|| || ||| ::  |   |:| :   |: || |||  : :|    |
Db        167 LENAYKAIRHGQCSAALVGGVNILLKPNTSVQFMKLGMLSPDGACKAFDVSGNGYCRSEA 226

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDG-ASNGLTAPNGPSQQRVIRQALANAGLS 298
              | ::|: : | |     || | :  :  | ||    |:| |:|  ||:::       |:
Db        227 VVVVLLTKKSMA----KRVYATIVNAGSNTDGFKEQGVTFPSGEMQQQLVGSLYRECGIK 282

Qy        299 AGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGV 358
              |||: |||||||| :||| |   ::  : |       || :|| |||:|| : |:|:| :
Db        283 PGDVEYVEAHGTGTKVGDPQEVNGIVNVFCQ--CEREPLLIGSTKSNMGHPEPASGLAAL 340

Qy        359 IKMVMALRHGLLPRTLHVDEPSPHVD-WSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVS 417
              |::::| |||    || ::|:| :     |::::: :  |  ||         ::|||
Db        341 AKVILSLEHGLWAPNLHFNDPNPDIPALHDGSLKVVCKPTPVKGG-----LVSINSFGFG 395

Qy        418 GTNAHVIL 425
              |:||||||
Db        396 GSNAHVIL 403


RESULT 13
STCA_EMENI
ID   STCA_EMENI     STANDARD;      PRT;  2181 AA.
AC   Q12397;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN   STCA OR PKSST.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
RX   MEDLINE=96202293; PubMed=8643646;
```

```
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 4;
RX   MEDLINE=95370159; PubMed=7642507;
RA   Yu J.-H., Leonard T.J.;
RT   "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT   novel type I polyketide synthase.";
RL   J. Bacteriol. 177:4792-4800(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34740; AAC49191.1; -.
DR   EMBL; L39121; AAA81586.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW   Multifunctional enzyme.
FT   DOMAIN      383    814       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      884   1209       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1706   1777       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1830   1901       ACYL CARRIER (ACP) 2.
FT   DOMAIN        ?   2181       THIOESTERASE.
FT   ACT_SITE    552    552       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    978    978       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1738   1738       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1862   1862       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2028   2028       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2181 AA;  238831 MW;  5A3E5712AA9AD942 CRC64;

  Query Match             31.0%;  Score 702.5;  DB 1;  Length 2181;
  Best Local Similarity   37.0%;  Pred. No. 1.5e-40;
  Matches  162;  Conservative   74;  Mismatches  185;  Indels   17;  Gaps    8;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||| |: |||     | | ||||:     |       |  | ||: |  ||         ||
Db        383 LAIVSMSGRFP-EAPSTDSFWDLLYKGLDVCKEVPL-RRWDVKTHVDPSGKARNKGATRW 440

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTP---TGV 119
              | :|  || ||  || |||:||  ||| ||: | : :| :|  || | |    |      ||
Db        441 GCWLDFAGEFDPRFFSISPKEAPQMDPAQRMALMSTYEAMERGGIVPDTTPSTQRNRIGV 500

Qy        120 FTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179
              | |    |:     |  |: | : :||     | |||:: | | ||: | ||||||| |
Db        501 FHGVTSNDW--METNTAQNIDTYFITGGNRGFIPGRINFCFEFSGPSYSNDTACSSSLAA 558

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239
              :|||| :|  |:|  |:|||  :: :|       :   |:  |:|||  ||||    ||
Db        559 IHLACNSLWRGDCDTAVAGGTNMIFTPDGHTGLDKGFFLSRTGNCKAFDDAADGYCRAEG 618

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299
              || : ::|| ||      :|| :      |     |: :| |  |:|   :   |: ||:|
Db        619 VGTVFIKRLEDALAENDPILATILDIKTNHSAMSDSMTRPFKPAQIDNMSALLSTAGISP 678

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQD---RAGEGPLWLGSVKSNVGHTQAAAGVA 356
              |:  :| ||||| :|| :| :::|: :  |   |  : ||::|| |:|:|| :  :||
Db        679 LDLSYIEMHGTGTQVGDAVEMESVLSLFAPDETFRPRDKPLYVGSAKANIGHGEGVSGVT 738

Qy        357 GVIKMVMALRHGLLPRTLHVDEPSPHV-----DWSAGAVQLLTETVPWPGGEGRLRRAGV 411
             . :||::: :::  :|    : :|    :       |  |  | :  |  |||  :   ||  :
Db        739 SLIKVLLMMKNDTIPPHCGI-KPGSRINRNYPDLPARNVHIAFEPKPWPRTD-TPRRVLI 796

Qy        412 SSFGVSGTNAHVILEEAP 429
              ::|  :| |  |::|:||
Db        797 NNFSAAGGNTAVLVEDAP 814


RESULT 14
PKS1_ASPPA
ID   PKS1_ASPPA     STANDARD;      PRT;  2109 AA.
AC   Q12053;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aflatoxin biosynthesis polyketide synthase (PKS).
GN   PKSL1.
OS   Aspergillus parasiticus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2999;
RX   MEDLINE=96042102; PubMed=7592391;
RA   Feng G.H., Leonard T.J.;
RT   "Characterization of the polyketide synthase gene (pksL1) required
RT   for aflatoxin biosynthesis in Aspergillus parasiticus.";
RL   J. Bacteriol. 177:6246-6254(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: Aflatoxin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L42766; AAC41675.1; -.
DR   EMBL; L42765; AAC41674.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Transferase; Acyltransferase; Phosphopantetheine;
KW   Multifunctional enzyme.
FT   DOMAIN      374    805       BETA-KETOACYL SYNTHASE.
FT   DOMAIN        ?      ?       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1714   1785       ACYL CARRIER (ACP).
```

```
FT   DOMAIN        ?   2109       THIOESTERASE.
FT   ACT_SITE    543    543       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    993    993       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1746   1746       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1937   1937       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2109 AA;  230715 MW;  CB701372A16D8551 CRC64;

  Query Match             30.2%;  Score 684.5;  DB 1;  Length 2109;
  Best Local Similarity   35.4%;  Pred. No. 2.5e-39;
  Matches  158;  Conservative   74;  Mismatches  181;  Indels   33;  Gaps    8;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||| |: ||| |: : |||| :     |     |  | ||::|  ||         |:
Db        374 LAIVSMSGRFPESPTT-ESFWDLLYKGLDVCKEVPR-RRWDINTHVDPSGKARNKGATKW 431

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTP---TGV 119
              | :|  :| ||  ||||||:||  ||| ||: | : :| :| ||: | |   |     ||
Db        432 GCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRDRIGV 491

Qy        120 FTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179
              | |    |:      |  |: | : :||     | |||:: | | ||: : ||||||| |
Db        492 FHGVTSNDW--METNTAQNIDTYFITGGNRGFIPGRINFCFEFAGPSYTNDTACSSSLAA 549

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239
              :|||| :|  :|  |:||| :: :|       :   |:  |:|| :   |||    ||
Db        550 IHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDKGFFLSRTGNCKPYDDKADGYCRAEG 609

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299
              || : ::|| ||  :   :| |:  :  |     |  :| |:  :|   :  ||   ||
Db        610 VGTVFIKRLEDALADNDPILGVILDAKTNHSAMSESMTRPHVGAQIDNMTAALNTTGLHP 669

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQD---RAGEGPLWLGSVKSNVGHTQAAAGVA 356
              |   :| ||||| :|| :| :::|: :      |  : ||::|| |:|||| :  :||
Db        670 NDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETARKADQPLFVGSAKANVGHGEGVSGVT 729

Qy        357 GVIKMVMALRHGLLPRTLHVDEPSPHV------------DWSAGAVQLLTETVPWPGGE 403
              :||::| ::|  :|         ||             |  |  | :  |  |||
Db        730 SLIKVLMMMQHDTIP---------PHCGIKPGSKINRNFPDLGARNVHIAFEPKPWPRTH 780

Qy        404 GRLRRAGVSSFGVSGTNAHVILEEAP 429
                ||  :::|  :| |  :|:|:||
Db        781 -TPRRVLINNFSAAGGNTALIVEDAP 805


RESULT 15
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
```

```
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match             30.1%;  Score 682;  DB 1;  Length 2505;
  Best Local Similarity   34.9%;  Pred. No. 4.5e-39;
  Matches  153;  Conservative   92;  Mismatches  159;  Indels   34;  Gaps   10;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCY- 59
              | : | ||: : |    :  :||       : |||        :| : | |       |
Db          2 EEVVIAGMSGKLPES-ENLQEFW------ANLIGG--------VDMVTDDDRRWKAGLYG 46

Qy         60 -TRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTG 118
               :  | | |   ||| |||: |::|  |||| ||||| ::| |   |||| :| || ||
Db         47 LPKRSGKLKDLSKFDASFFGVHPKQAHTMDPQLRLLLEVSYEAIVDGGINPASLRGTNTG 106

Qy        119 VFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLV 178
              |: | :| : :  :    ::  |::: |   :::: |:|: | |:||:::|||||||:
Db        107 VWVGVSGSEASEALSRDPETLLGYSMVGCQRAMMANRLSFFFDFKGPSIALDTACSSSLL 166

Qy        179 ALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGE 238
              ||  | ||:|:|||  |: ||: ::   |   |:| :   |: || |::|  : :|    |
Db        167 ALQNAYQAIRSGECPAAIVGGINLLLKPNTSVQFMKLGMLSPDGTCRSFDDSGNGYCRAE 226

Qy        239 GVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGA-SNGLTAPNGPSQQRVIRQALANAGL 297
              |  :|: : | |      || | :  :  | ||    |:| |:| :|:::||      |:
Db        227 AVVAVLLTKKSLA----RRVYATILNAGTNTDGCKEQGVTFPSGEAQEQLIRSLYQPGGV 282

Qy        298 SAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAG 357
              :   :: :|||||| :||| |   :  :    |  : || :|| |||:|| : |:|:|
Db        283 APESLEYIEAHGTGTKVGDPQELNGITRSLCAFR--QSPLLIGSTKSNMGHPEPASGLAA 340

Qy        358 VIKMVMALRHGLLPRTLHVDEPSPHVD-WSAGAVQLLTETVPWPGGEGRLRRAGVSSFGV 416
              : |:::| :|:    ||   |:| :     | :|::.  :|  ||       |::|||
Db        341 LTKVLLSLENGVWAPNLHFHNPNPEIPALLDGRLQVVDRPLPVRGG-----IVGINSFGF 395

Qy        417 SGTNAHVIL----EEAPA 430
              | | ||||    ::|||
Db        396 GGANVHVILQPNTQQAPA 413


Search completed: June 17, 2003, 13:02:37

Job time : 5.68929 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 9.60187 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_2086_2515
Perfect score: 2265
Sequence: 1 EPIAIVGMACRFPGGVTSAD.....VSSFGVSGTNAHVILEEAPA 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1654	73.0	3519	2	S43048	polyketide synthas
2	1638	72.3	2723	2	T03221	probable polyketid
3	1636	72.2	5069	2	T17464	rifamycin polyketi
4	1628.5	71.9	3739	2	T17410	polyketide synthas
5	1622.5	71.6	4613	2	T17409	polyketide synthas
6	1611	71.1	4735	2	T17463	rifamycin polyketi
7	1591.5	70.3	3413	2	T17467	rifamycin polyketi
8	1590	70.2	7576	2	T17428	FK506 polyketide s
9	1586	70.0	10223	2	T30225	polyketide synthas
10	1581.5	69.8	8563	2	T30226	polyketide synthas
11	1574	69.5	2100	2	T03223	probable polyketid
12	1570.5	69.3	1728	2	T17466	rifamycin polyketi
13	1567.5	69.2	1763	2	T17465	rifamycin polyketi
14	1563	69.0	1937	2	T03224	probable polyketid
15	1561.5	68.9	1562	2	T17411	polyketide synthas
16	1559	68.8	6260	2	T30228	polyketide synthas
17	1558.5	68.8	4151	2	G70944	probable polyketid
18	1556	68.7	2103	2	G86925	probable polyketid
19	1554	68.6	6420	2	T30283	polyketide synthas
20	1550	68.4	496	2	H70668	probable polyketid
21	1544.5	68.2	3573	2	S23070	erythronolide synt
22	1538	67.9	1346	2	T17412	polyketide synthas
23	1533.5	67.7	1762	2	T03222	probable polyketid
24	1521	67.2	1602	2	H70984	probable polyketid
25	1497.5	66.1	2126	2	H70621	probable polyketid
26	1495.5	66.0	3491	2	T43231	probable 6-deoxyer
27	1425.5	62.9	3172	2	S22012	erythronolide synt
28	1425.5	62.9	3178	2	S13595	6-deoxyerythronoli
29	1170.5	51.7	2124	2	T28658	polyketide synthas
30	1164	51.4	2201	2	S73014	polyketide synthas
31	1163.5	51.4	2518	2	AI2140	polyketide synthas
32	1151.5	50.8	2188	2	A70984	probable polyketid
33	1148	50.7	1587	2	AB2012	hypothetical prote
34	1139.5	50.3	1827	2	B70984	probable polyketid
35	1135.5	50.1	1815	2	S73021	polyketide synthas
36	1135.5	50.1	1822	2	F87203	polyketide synthas
37	1110	49.0	1446	2	S73013	polyketide synthas
38	1110	49.0	1540	2	H87203	polyketide synthas
39	1109.5	49.0	444	2	A70674	probable mbtC prot
40	1104.5	48.8	1538	2	E70874	probable ppsB prot
41	1101	48.6	1293	2	T30871	orsellinic acid sy
42	1095	48.3	2478	2	AH2140	polyketide synthas
43	1085.5	47.9	1774	2	T17421	polyketide synthas
44	1066	47.1	1876	2	C70749	probable ppsA prot
45	1054	46.5	1871	2	A87204	polyketide synthas

ALIGNMENTS

RESULT 1
S43048
polyketide synthase type I - Streptomyces antibioticus
N;Contains: acyl carrier protein; acyltransferase; ketoacylsynthase; ketoreductase; thi
C;Species: Streptomyces antibioticus
C;Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 26-May-2000
C;Accession: S43048; S41729
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, February 1993
A;Reference number: S43048
A;Accession: S43048
A;Molecule type: DNA
A;Residues: 1-3519 <SWA>
A;Cross-references: EMBL:L09654; NID:g153407; PIDN:AAA19695.1; PID:g153408
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A;Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyket
A;Reference number: S41729; MUID:94150470; PMID:8107683
A;Accession: S41729
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1683-3238,3273-3303,'K',3305-3407,'T',3409-3462,'Y',3464-3516,'E',3518-3519
A;Cross-references: EMBL:L09654
C;Genetics:
A;Start codon: GTG
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;570-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1202-1381/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1489-1560/Domain: acyl carrier protein homology <ACP1>
F;1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3143-3214/Domain: acyl carrier protein homology <ACP2>
F;3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 73.0%; Score 1654; DB 2; Length 3519;
Best Local Similarity 73.3%; Pred. No. 5.8e-106;
Matches 315; Conservative 40; Mismatches 75; Indels 0; Gaps 0;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              | |||  |:||| ||: | :| |  ::  :||: | | ||||||| || |||::||| |
Db       1687 EAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGTTYV 1746

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | | |||  ||| |||||||||||||||||||||:||  | | |:|:|: || ||:|
Db       1747 REGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTATGIF 1806

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |   | |      | :|  |: ||||| :|:|||||||| |||||:||||||||||||
```

```
Db       1807 IGAGHQGYGPDPKRAPESVAGYLLTGTASAVLSGRISYTFGLEGPAVTVDTACSSSLVAL 1866

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240

Db       1867 HLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAADGMGWGEGV 1926

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300

Db       1927 SLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 1986

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360

Db       1987 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAAAGVAGVMK 2046

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420

Db       2047 MVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGISAFGVSGTN 2106

Qy        421 AHVILEEAPA 430

Db       2107 AHVIVEEAPA 2116


RESULT 2
T03221
probable polyketide synthase module 1 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 17-Nov-2000
C;Accession: T03221
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03221
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2723 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38061.1; PID:g2624948
A;Experimental source: ATCC 29253
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-504/Domain: acetate-CoA ligase homology <ACL>
F;881-949/Domain: acyl carrier protein homology <ACP1>
F;996-1397/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;1509-1794/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;2568-2639/Domain: acyl carrier protein homology <ACP2>

  Query Match             72.3%;  Score 1638;  DB 2;  Length 2723;
  Best Local Similarity   71.3%;  Pred. No. 5.4e-105;
  Matches  306;  Conservative   49;  Mismatches   72;  Indels    2;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60

Db        975 EPIAIVGMACRLPGGVTSPEELWELLSSGGDGISGFPTDRGWDLEHLYHPDPDHPGTSYV 1034

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120

Db       1035 REGGFLHDAGDFDAALFGISPREALAMDPQQRLLLETTWEVLERAGIDPTSLRGSRTGVF 1094

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180

Db       1095 AGVMYHDYISRLRTVPPELEGYLSSGGSGSVVSGRVSYTFGFEGPAVSVDTACSSSLVAL 1154

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240

Db       1155 HLAVQALRTGECTLAVAGGVTVMASPSSYIEFSRQRGLSPDGRCKSFAADADGAAWSEGA 1214

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300

Db       1215 ATLLVERLSDAQRLGHPVLGVVRGTAVNQDGASNGLTAPNGPSQQRVIRDALTNAQLTAT 1274

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360

Db       1275 DIDLIEAHGTGTKLGDPIEAQAVLAMYGQDR--DLPVWLGSLKSNIGHTQAAAGVAGVMK 1332

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420

Db       1333 AVLSMRHGVMPKSLHIDEPTPEVDWSAGTVELLAEARDWPGADERPRRAGVSSFGVSGTN 1392

Qy        421 AHVILEEAP 429

Db       1393 AHVILEQAP 1401


RESULT 3
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             72.2%;  Score 1636;  DB 2;  Length 5069;
  Best Local Similarity   72.4%;  Pred. No. 1.6e-104;
  Matches  315;  Conservative   43;  Mismatches   65;  Indels   12;  Gaps    4;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60

Db       3332 EPIAIVGMACRLPGGVAGPEDLWRLVAEGRDAVSSFPTDRGWDTDSLYDPDPARPGKTYT 3391

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120

Db       3392 RHGGFLHEAGLFDAGFFGISPREAVAMDPQQRLLLEASWEAMEDAGVDPLSLKGNDVGVF 3451

Qy        121 TGTNGQDYALRVHNAGQST-----DGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSS 175

Db       3452 TGMFGQGYVA----PGDSVVTPELEGFAGTGGSSSVASGRVSYVFGFEGPAVTIDSACSS 3507

Qy        176 SLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTG 235

Db       3508 SLVAMHLAAQSLRQGECSMALAGGATVMANPGAFVEFSRQRGLAVDGRCKAFAAAADGTG 3567

Qy        236 WGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANA 295

Db       3568 WAEGVGVVILERLSVARERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRRALVSA 3627

Qy        296 GLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGV 355

Db       3628 GLAPSDVDVVEAHGTGTTLGDPIEAQALLATYGKDR--ESPLWLGSLKSNIGHAQAAAGV 3685

Qy        356 AGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFG 415

Db       3686 AGVIKMVQALRHEVLPPTLHVDRPTPEVDWSAGAVELLTEAREWP-RNGRPRRAGVSAFG 3744

Qy        416 VSGTNAHVILEEAPA 430

Db       3745 VSGTNAHLILEEAPA 3759


RESULT 4
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
```

```
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

  Query Match             71.9%;  Score 1628.5;  DB 2;  Length 3739;
  Best Local Similarity   72.6%;  Pred. No. 3.6e-104;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||||||||| |||| | :| | |::  :||| ||| |||||:: |||||||   |  |
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              .  ||||  |   : |  :  :|:  || | |::|||:||| | |||||:|||||||||||
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :|||||||||:|  ||||||||||| ||  |||:|:||| |  |||
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||||| |||||||||||||||||||||||||||||||||:|||:| |:
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:||||| |  | || |||:|||:|||||||:|:||
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:|||| ||  :||||| |:|||| : ||   || |||| |||||:|||
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 5
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

  Query Match             71.6%;  Score 1622.5;  DB 2;  Length 4613;
  Best Local Similarity   72.1%;  Pred. No. 1.2e-103;
  Matches  310;  Conservative   42;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||||||||| |||| | :| | |::  :||| ||| |||||:: |||||||   |  |
Db       1109 EPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYV 1168

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||: :   ||| ||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1169 RQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVF 1228

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    :|   : : |:  ||: |||   ||:|||:||| | ||||::|||||||||||
Db       1229 TGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVAL 1288

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||  ||||||| || :|| |||||||||| ||  |||:|:|||| | |||
Db       1289 HLAVQALRKGEVDMALAGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGV 1348

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||:|||||||||:|||||||||||||||||||||:|||:| |:
Db       1349 GVLLVERLSDARRNGHQVLAVVRGSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTS 1408

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||||  |  || |||:|||:|||||||||:||||
Db       1409 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIK 1468

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||||:||||||||  :|||||||:|||| | ||   :| |||| |||||:|||
Db       1469 MVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGT 1528

Qy        420 NAHVILEEAP 429
              ||||:|||||
Db       1529 NAHVVLEEAP 1538


RESULT 6
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

  Query Match             71.1%;  Score 1611;  DB 2;  Length 4735;
  Best Local Similarity   73.3%;  Pred. No. 7.9e-103;
  Matches  315;  Conservative   33;  Mismatches   78;  Indels    4;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||| |||||||| | :| | |::   ||:  || ||||| | ||  |||| || |
Db       2195 EPIAIVAMACRFPGGVHSPEDLWRLVADGADAVTEFPADRGWDTDRLYHEDPDHEGTTYV 2254
```

```
Qy          61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db        2255 RHGAFLDDAAGFDAAFFGISPNEALAMDPQQRLLLETSWELFERAAIDPTTLAGQDIGVF 2314

Qy         121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db        2315 AGVNSHDYSMRMHRAA-GVEGFRLTGGSASVLSGRVAYHFGVEGPAVTVDTACSSSLVAL 2373

Qy         181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db        2374 HMAVQALQRGECSMALAGGVMVMGTVETFVEFSRQRGLAPDGRCKAFADGADGTGWSEGV 2433

Qy         241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db        2434 GLLLVERLSEAQRRGHQVLAVVRGSAVNSDGASNGLTAPNGPSQQRVIRKALAAAGLSTS 2493

Qy         301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db        2494 DVDAVEAHGTGTTLGDPIEAEALLATYGQNR--ETPLWLGSVKSNLGHTQAAAGVAGVIK 2551

Qy         361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db        2552 MVMAMRHGVLPRTLHVDRPSSYVDWSAGAVELLTEARDWV-SNGHPRRAGVSSFGIGGTN 2610

Qy         421 AHVILEEAPA 430
Db        2611 AHVVLEEVAA 2620
```

RESULT 7
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

Query Match 70.3%; Score 1591.5; DB 2; Length 3413;
Best Local Similarity 72.0%; Pred. No. 1.2e-101;
Matches 311; Conservative 37; Mismatches 73; Indels 11; Gaps 4;

```
Qy           1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db          31 EPLAVVGMACRFPGGVSSPEDLWQLVAGGVDALSDFPDDRGWELDGLFDPDPDHPGTSYT 90

Qy          61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db          91 SQGGFLRGAGLFDAGLFGISPREALVMDPQQRVLLETSWEALEDAGVDPLSLKGSDVGVF 150

Qy         121 TGTNGQDYALRVHNAGQST---DGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSL 177
Db         151 SGVFTQGY-----GAGAITPDLEAFAGIGAASSVASGRVSYVFGLEGPAVTIDTACSSSL 205

Qy         178 VALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWG 237
Db         206 VAIHLAAQALRAGECSMALAGGATVMPTPGTFVAFSRQRVLAADGRSKAFSSTADGTGWA 265

Qy         238 EGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGL 297
Db         266 EGAGVLVLERLSVAQERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRKALAGAGL 325

Qy         298 SAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAG 357
Db         326 VASDVDVVEAHGTGTALGDPIEAQALLATYGQGR--ERPLWLGSVKSNFGHTQAAAGVAG 383

Qy         358 VIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVS 417
Db         384 VIKMVQALRHGAMPPTLHVAEPTPEVDWSAGAVELLTEPREWPAGD-RPRRAGVSAFGIS 442

Qy         418 GTNAHVILEEAP 429
Db         443 GTNAHLILEEAP 454
```

RESULT 8
T17428
FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: T17428
R;Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressar
A;Reference number: Z18779; MUID:98451508; PMID:9780228
A;Accession: T17428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7576 <MOT>
A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A;Experimental source: strain MA6548
C;Genetics:
A;Gene: fkbB
C;Function:
A;Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-500/Domain: acetate-CoA ligase homology <ACL>
F;1095-1166/Domain: acyl carrier protein homology <ACP1>
F;1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;2680-2751/Domain: acyl carrier protein homology <ACP2>
F;2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;4320-4391/Domain: acyl carrier protein homology <ACP3>
F;4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5903-5974/Domain: acyl carrier protein homology <ACP4>
F;6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;7450-7521/Domain: acyl carrier protein homology <ACP5>

Query Match 70.2%; Score 1590; DB 2; Length 7576;
Best Local Similarity 73.9%; Pred. No. 4e-101;
Matches 315; Conservative 37; Mismatches 66; Indels 8; Gaps 3;

```
Qy           1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db        1183 EPVAIVGMACRLPGGVTSPEGLWRLVASGTDAITEFPVDRGWDVDELVDPDPDAVGRSTT 1242

Qy          61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db        1243 RHGGFLTEATGFDAAFFGISPNEALAMDPQQRLVLETSWEAFEHAGIVPDTLRESDTGVF 1302

Qy         121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db        1303 MGAFHQGY-----GAGRDLGGLGVTATQTSVLSGRLSYFYGLQGPAVTVDTACSSSLVAL 1357

Qy         181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db        1358 HQAAQALRSGECSLALAGGVTVMATPGSFVEFSRQRGLSPDGRCKAFADAADGTGFAEGV 1417

Qy         241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db        1418 GVLVVERQSDAERNGHTVLAVVRGSAVNQDGASNGLSAPNGVAQQRVIRQALANAGLSGA 1477
```

```
Qy       301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db      1478 DVDAVEAHGTGTVLGDPIEAQAVLASYGQER--EVPLLLGSVKSNVGHTQAAAGVAGVIK 1535

Qy       361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db      1536 MVMAMRHGLLPKTLHVDEPSSHVDWSAGAVEVVTETRPWP-ESGGARRAGVSSFGVSGTN 1594

Qy       421 AHVILE 426
Db      1595 AHVILE 1600
```

RESULT 9
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

Query Match 70.0%; Score 1586; DB 2; Length 10223;
Best Local Similarity 72.7%; Pred. No. 1.1e-100;
Matches 312; Conservative 34; Mismatches 75; Indels 8; Gaps 3;

```
Qy         1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db        33 EPLAIVGMACRLPGGVSSPEDLWRLVESGTDAISGFPADRGWDVENLYDPDPEAAGKSYC 92

Qy        61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db        93 VQGGFLDSAGGFDASFFGISPREALAMDPQQRLVLEASWEAFERAGIEPGSLRGSDTGVF 152

Qy       121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       153 MGAYPGGYGV-----GADLGGFGATAGAVSVLSGRVSYFFGLEGPAVTVDTACSSSLVAL 207

Qy       181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       208 HQAAYALRQGECSLALVGGVTVMATPQSFVEFSRQRGLASDGRCKAFADSADGTGWAEGA 267

Qy       241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       268 GVLLVERLSDAQRKGHQVLAVVRGSAVNQDGASNGLSAPNGPSQQRVIRAALSNAGLSTA 327

Qy       301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       328 EVDVVEAHGTGTTLGDPIEAQALLATYGQDR--EQPLLLGSVKSNLGHTQAAAGVSGVIK 385

Qy       361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db       386 MVMALQHVLVPRTLHVDEPSRHVDWTDGAVALVTENQPWP-DMGRPRRAGVSSFGISGTN 444

Qy       421 AHVILEEAP 429
Db       445 AHVILESAP 453
```

RESULT 10
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Stau
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 69.8%; Score 1581.5; DB 2; Length 8563;
Best Local Similarity 72.0%; Pred. No. 1.8e-100;
Matches 309; Conservative 40; Mismatches 71; Indels 9; Gaps 4;

```
Qy         1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db      6854 EPLAIVGMACRLPGGVSSPEDLWRLVESGTDAVSGFPTDRGWDVENLYDSDPEAAGKSYC 6913

Qy        61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db      6914 VQGGFLDTAAGFDAGFFGISPREALAMDPQQRLLLEVSWEAFERAGIEPGSVRGSDTGVF 6973

Qy       121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db      6974 IGAFPVGY-----GAGFDREGYGAT-SGPSVLSGRVSYVFGLEGPAITMDTACSSSLVAL 7027

Qy       181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db      7028 HLAAQALRNGECSMALAGGVTVMATPEVFTEFARQRGLASDGRCKAFADSADGAGFSEGA 7087

Qy       241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db      7088 GLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALSNAGLSTA 7147

Qy       301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db      7148 DVDVVEAHGTGTTLGDPIEAQALLATYGQDR--EQPLLLGSLKSNIGHTQAASGVSGVIK 7205
```

```
Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              |||||||| :|||||||||| |||||:||||:|: |  |||| : | ||||||||||||||
Db       7206 MVMALRHGFVPRTLHVDEPSRHVDWAAGAVELVRENQPWPGTD-RPRRAGVSSFGVSGTN 7264

Qy        421 AHVILEEAP 429
              |||:|| ||
Db       7265 AHVVLESAP 7273


RESULT 11
T03223
probable polyketide synthase module 3 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03223
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2100 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38063.1; PID:g2624950
A;Experimental source: ATCC 29253
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;566-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1952-2027/Domain: acyl carrier protein homology <ACP>

  Query Match             69.5%;  Score 1574;  DB 2;  Length 2100;
  Best Local Similarity   69.5%;  Pred. No. 1.1e-100;
  Matches  298;  Conservative   50;  Mismatches   79;  Indels    2;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||:|||||||||||| :  |:|:||  ||| ||| |||||::  :| ||   || |
Db         34 EPIAIIGMACRFPGGVTSPEALWELVSSGTDAISGFPDDRGWDMEGFFDSDPAQEGTSYM 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|   ||  |||||||||:|||||||||||| || :| |||:| :| |: ||||
Db         94 REGGFLTNATGFDPAFFGISPREAVAMDPQQRLLLETTWEVLERAGIDPTSLRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    |  ||       : :|  | | || |||:||||||||:|||||||||| ||
Db        154 TGLMANSYTTRVRRIPGELESYAGNGEAASVASGRVSYTFGFEGPAMSVDTACSSSLTAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| :|||:||| :|:|||||| ::|  | : ||||||| || ||||| :|||||  |||
Db        214 HLAVRALRSGECPLAVAGGVTVFATPRIFSDLSRQRGLARDGRCKAFSTSADGTGLSEGV 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| | || || ||||:||||||||||||||||:|||||: |||:| |:|
Db        274 GLLLVERLSDAQRLGHSVLGVVRGTAVNQDGASNGLTAPNGPAQQRVIKAALADARLTAA 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |:| :||||||| |||||||::|||||||  | |||:||:|||:||:|||||||:||
Db        334 DIDLIEAHGTGTKLGDPIEAQAIIATYGQDR--EQPLWMGSLKSNIGHSQAAAGVAGIIK 391

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              :|||:|||::|:|||:|||:| |||||| |:|| |   ||| : | ||||||||||||||
Db        392 IVMAMRHGVMPKTLHIDEPTPEVDWSAGTVELLAEARDWPGVDERPRRAGVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              ||||||:||
Db        452 AHVILEQAP 460


RESULT 12
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17466
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17466
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1728 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1573-1644/Domain: acyl carrier protein homology <ACP>

  Query Match             69.3%;  Score 1570.5;  DB 2;  Length 1728;
  Best Local Similarity   71.6%;  Pred. No. 1.4e-100;
  Matches  307;  Conservative   31;  Mismatches   90;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| ||||:||||:| :|  | |::   ||:  || |||||:| | ||||| ||| ||
Db         34 EPIAIVAMACRYPGGVSSPEDLWRLVAEGTDAVSAFPGDRGWDVDGLVDPDPDRPGTTYT 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
                ||||::|| ||| |||||||||:||||||||||||:|| ||  | :| :| |:  |||
Db         94 DQGGFLHEAGLFDAGFFGISPREAVAMDPQQRLLLETSWEAIERTGTDPLSLKGSDIGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    |           :||  || |  : |||:||  |||||||:||| |||||||:
Db        154 TGVASMGYGAGGGVVAPELEGFVGTGAAPCIASGRVSYVLGFEGPAVTVDTGCSSSLVAM 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||||||||||  ||: ||:|| |||||||| || ||||| :||| |  |||
Db        214 HLAAQALRRGECSMALAGGAMVMAQPGSFVSFSRQRGLALDGRCKAFSDSADGMGLAEGV 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:: :|||| |   |||||||:|| |||||||||||||||||||||||| ||| ||||
Db        274 GVIALERLSVARERGHRVLAVLRGIAVNQDGASNGLTAPNGPSQQRVIRAALAEAGLSPS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||| |||||||||||||||||||: |  | ||||||||||:|||||||| |||
Db        334 DVDAVEGHGTGTTLGDPIEAQALLATYGKGRDPEKPLWLGSVKSNLGHTQAAAGVASVIK 393

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              || |||||:|| ||||| ||  |||||||| ||||   ||  ||| |||||||||:||||
Db        394 MVQALRHGVLPPTLHVDRPSTEVDWSAGAVSLLTEAREWP-REGRPRRAGVSSFGISGTN 452

Qy        421 AHVILEEAP 429
              ||:||||||
Db        453 AHLILEEAP 461


RESULT 13
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>
```

```
  Query Match           69.2%;  Score 1567.5;  DB 2;  Length 1763;
  Best Local Similarity  71.1%;  Pred. No. 2.4e-100;
  Matches  305;  Conservative   37;  Mismatches   82;  Indels    5;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||| |||||   | | |::   ||: |||||| ||||||:|||||| || ||
Db         36 EPIAIVGMACRLPGGVTGPGDLWRLVAEGGDAVSGFPTDRCWDLDTLFDPDPDHAGTSYT 95

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||:||  ||  |||||||||||||||||||| :|| :|  |::| :| || |||
Db         96 DQGGFLHDAALFDPGFFGISPREALAMDPQQRLLLEASWEALEGVGLDPASLQGTDVGVF 155

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||  |  |   :  |      || || | || |||:|| ||||||||::|||||||||:
Db        156 TGAGGSGYGGGL--TGPEMQSFAGTGLASSVASGRVSYVFGFEGPAVTIDTACSSSLVAM 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| |:|||||||  ||| | :|| |||||||| || ||||:: |||     ||:
Db        214 HLAAQALRQGDCSMALAGGAMVMSGPDSFVVFSRQRGLATDGRCKAFASGADGMVLAEGI 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ::::|||| |   |||||||:|||||||||||||||||||||||||| ||||||:
Db        274 SVVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQRVIRAALANAGIGPS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| ||||||||:|||||||||||||||||  | ||||||:|||:|||||||||| |||
Db        334 DVDLVEAHGTGTSLGDPIEAQALLATYGQDR--ETPLWLGSLKSNIGHTQAAAGVASVIK 391

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              :| |||||::| ||||||||  |||| |||:||| :  || |: | ||||||||||||||
Db        392 VVQALRHGVMPPTLHVDEPSSQVDWSEGAVELLTGSRDWPRGD-RPRRAGVSSFGVSGTN 450

Qy        421 AHVILEEAP 429
              |:|:||||
Db        451 VHLIIEEAP 459


RESULT 14
T03224
probable polyketide synthase module 4 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03224
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03224
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1937 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38064.1; PID:g2624951
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;549-827/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

  Query Match           69.0%;  Score 1563;  DB 2;  Length 1937;
  Best Local Similarity  69.5%;  Pred. No. 5.5e-100;
  Matches  298;  Conservative   49;  Mismatches   80;  Indels    2;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| |||| |||| | :  |||::: :| :  || ||||||: :|||||||||| |
Db         34 EPIAIVSMACRLPGGVDSPESLWDLVAAGRDGVSAFPIDRGWDLEGIYDPDPDHPGTTYA 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:|||| ||  ||  |||||||||||||||||||| || :| |||:|  | |: ||||
Db         94 RDGGFLDDAAGFDPAFFGISPREALAMDPQQRLLLETTWEVLERAGIDPAALRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||     |||    :     :|:   |:| || |||:|||||||||||:||||||||||
Db        154 TGVIHNDYADWSADDTDQLEGYLGNGSAASVASGRVSYTFGFEGPAVSLDTACSSSLVAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| ||||:|||::|: ||||||:|   |:|||||||| || ||||:| |:| |  ||:
Db        214 HLAIQALRSGECTLAITGGVTVLASVAPFIEFSRQRGLADDGRCKAFAADANGFGMAEGI 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| | || |||||||:|||||||||||||||||||||||| || || |:|
Db        274 GLLLVERLSDAQRLGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRDALTNAQLTAT 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |:| |||||||| |||||||||::|||| |  : |: |||:|||:|| ||||||||:||
Db        334 DIDLVEAHGTGTKLGDPIEAQAIIATYGQHR--DQPVRLGSLKSNIGHAQAAAGVAGIIK 391

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              |||:|||::|||||:|||:| |||: | |:||||:  ||  :   ||| ||||||||||
Db        392 TVMAMRHGIMPRTLHIDEPTPEVDWTTGTVELLTESRDWPTTDQHPRRAAVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              |||||:||
Db        452 AHVILEQAP 460


RESULT 15
T17411
polyketide synthase III - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17411
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17411
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1562 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C;Genetics:
A;Gene: pikAIII
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1403-1474/Domain: acyl carrier protein homology <ACP>

  Query Match           68.9%;  Score 1561.5;  DB 2;  Length 1562;
  Best Local Similarity  69.8%;  Pred. No. 5.4e-100;
  Matches  305;  Conservative   47;  Mismatches   74;  Indels   11;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |:: : |||  || |||||:: |||||||  |  |
Db         35 EPVAIVGMACRLPGGVASPEDLWQLVAGDGDAISEFPQDRGWDVEGLYDPDPDASGRTYC 94

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| |||:|||||||||||||||||| | |||| || |||:|  | |:  |||
Db         95 RSGGFLHDAGEFDADFFGISPREALAMDPQQRLSLTTAWEAIESAGIDPTALKGSGLGVF 154

Qy        121 TG------TNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACS 174
              |       |:||  |::        :|  ::| |   :||||:|  | :|||::||||||
Db        155 VGGWHTGYTSGQTTAVQ----SPELEGHLVSGAALGFLSGRIAYVLGTDGPALTVDTACS 210

Qy        175 SSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGT 234
              ||||||||| |||| ||| |||||||||| :   ||:||||||||||  |||: :|||
Db        211 SSLVALHLAVQALRKGECDMALAGGVTVMPNADLFVQFSRQRGLAADGRSKAFATSADGF 270

Qy        235 GWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALAN 294
              |  || |:|||||||| |||||:|||||||||||||||||||||:||||||||||:|||:
Db        271 GPAEGAGVLLVERLSDARRNGHRILAVVRGSAVNQDGASNGLTAPHGPSQQRVIRRALAD 330

Qy        295 AGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAG 354
              | |: |||| |||||||| ||||||||||:|||||::: | || ||::|||:||||||||
Db        331 ARLAPGDVDVVEAHGTGTRLGDPIEAQALIATYGQEKSSEQPLRLGALKSNIGHTQAAAG 390
```

```
Qy        355 VAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSS 413
              ||||||||  |:||||||:||||||||  :|||||  |:||||  |  ||    :|  ||||  |||
Db        391 VAGVIKMVQAMRHGLLPKTLHVDEPSDQIDWSAGTVELLTEAVDWPEKQDGGLRRAAVSS 450

Qy        414 FGVSGTNAHVILEEAPA 430
              ||:|||||||:||||||
Db        451 FGISGTNAHVVLEEAPA 467
```

Search completed: June 17, 2003, 13:12:23
Job time : 10.6019 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 13.398 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_2086_2515
Perfect score: 2265
Sequence: 1 EPIAIVGMACRFPGGVTSAD.....VSSFGVSGTNAHVILEEAPA 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1654	73.0	3519	9	US-09-808-880-4	Sequence 4, Appli
2	1646.5	72.7	4150	9	US-09-808-880-2	Sequence 2, Appli
3	1628.5	71.9	3739	9	US-09-860-846-33	Sequence 33, Appl
4	1628.5	71.9	3739	9	US-09-988-384B-33	Sequence 33, Appl
5	1628.5	71.9	3739	9	US-09-836-821-33	Sequence 33, Appl
6	1628.5	71.9	3739	9	US-09-793-708-2	Sequence 2, Appli
7	1628.5	71.9	3739	10	US-09-861-289-33	Sequence 33, Appl
8	1628.5	71.9	11877	9	US-09-860-846-6	Sequence 6, Appli
9	1628.5	71.9	11877	9	US-09-836-821-6	Sequence 6, Appli
10	1628.5	71.9	11877	10	US-09-861-289-6	Sequence 6, Appli
11	1628.5	71.9	12199	9	US-09-988-384B-6	Sequence 6, Appli
12	1625.5	71.8	4551	9	US-09-793-708-1	Sequence 1, Appli
13	1622.5	71.6	4613	9	US-09-860-846-31	Sequence 31, Appl
14	1622.5	71.6	4613	9	US-09-988-384B-31	Sequence 31, Appl
15	1622.5	71.6	4613	9	US-09-836-821-31	Sequence 31, Appl
16	1622.5	71.6	4613	10	US-09-861-289-31	Sequence 31, Appl
17	1609	71.0	3816	9	US-09-808-880-3	Sequence 3, Appli
18	1570	69.3	5215	9	US-09-860-846-2	Sequence 2, Appli
19	1570	69.3	5215	9	US-09-988-384B-2	Sequence 2, Appli
20	1570	69.3	5215	9	US-09-836-821-2	Sequence 2, Appli
21	1570	69.3	5215	10	US-09-861-289-2	Sequence 2, Appli
22	1561.5	68.9	1562	9	US-09-860-846-35	Sequence 35, Appl
23	1561.5	68.9	1562	9	US-09-988-384B-35	Sequence 35, Appl
24	1561.5	68.9	1562	9	US-09-836-821-35	Sequence 35, Appl
25	1561.5	68.9	1562	9	US-09-793-708-3	Sequence 3, Appli
26	1561.5	68.9	1562	10	US-09-861-289-35	Sequence 35, Appl
27	1543	68.1	1346	9	US-09-793-708-4	Sequence 4, Appli
28	1538	67.9	1346	9	US-09-860-846-37	Sequence 37, Appl
29	1538	67.9	1346	9	US-09-988-384B-37	Sequence 37, Appl
30	1538	67.9	1346	9	US-09-836-821-37	Sequence 37, Appl
31	1538	67.9	1346	10	US-09-861-289-37	Sequence 37, Appl
32	1382.5	61.0	7257	9	US-10-014-717-5	Sequence 5, Appli
33	1293	57.1	458	10	US-09-892-870-4	Sequence 4, Appli
34	1180.5	52.1	3798	9	US-10-014-717-6	Sequence 6, Appli
35	1178	52.0	1421	9	US-10-014-717-2	Sequence 2, Appli
36	1139.5	50.3	1827	9	US-09-712-363-261	Sequence 261, App
37	1085.5	47.9	2439	9	US-10-014-717-7	Sequence 7, Appli
38	977.5	43.2	1402	9	US-09-712-363-166	Sequence 166, App
39	942	41.6	1610	9	US-09-738-626-6666	Sequence 6666, Ap
40	887	39.2	1812	10	US-09-775-938A-38	Sequence 38, Appl
41	860	38.0	1832	9	US-10-014-717-4	Sequence 4, Appli
42	848	37.4	3032	9	US-09-836-705-44	Sequence 44, Appl
43	783	34.6	2563	9	US-09-836-705-46	Sequence 46, Appl
44	715	31.6	2509	9	US-10-237-271-1	Sequence 1, Appli
45	665	29.4	2756	9	US-10-331-061-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-808-880-4
; Sequence 4, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3519
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-808-880-4

Query Match 73.0%; Score 1654; DB 9; Length 3519;
Best Local Similarity 73.3%; Pred. No. 2.2e-121;
Matches 315; Conservative 40; Mismatches 75; Indels 0; Gaps 0;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              | |||  |:||| ||: | :| |  ::  :||: | | ||||||| || |||::||| |
Db       1687 EAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGTTYV 1746

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | |||  ||| |||||||||||||||||||||||:||  | | |:|:|: || ||:|
Db       1747 REGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTATGIF 1806
```

```
Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       1807 IGAGHQGYGPDPKRAPESVAGYLLTGTASAVLSGRISYTFGLEGPAVTVDTACSSSLVAL 1866

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       1867 HLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAADGMGWGEGV 1926

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       1927 SLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 1986

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       1987 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAAAGVAGVMK 2046

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db       2047 MVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGISAFGVSGTN 2106

Qy        421 AHVILEEAPA 430
Db       2107 AHVIVEEAPA 2116
```

RESULT 2
US-09-808-880-2
; Sequence 2, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-808-880-2

Query Match 72.7%; Score 1646.5; DB 9; Length 4150;
Best Local Similarity 72.8%; Pred. No. 1.1e-120;
Matches 313; Conservative 41; Mismatches 75; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db       2546 EPIAIVGMACRFPGGIASPEDLWDVLTAGEDVVSGLPQNRGWDLGRLYDPDPDRAGTSYM 2605

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db       2606 REGAFLHEAGEFDAAFFGISPREALAMDPQQRLLLETSWEALERAGITPSKLAGSPTGVF 2665

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       2666 FGMSNQDYAAQAGDVPSELEGYLLTGSISSVASGRVAYTFGLEGPAVTVDTACSSSLVAL 2725

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       2726 HLAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCKAFAASADGFGAAEGV 2785

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       2786 GVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 2845

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       2846 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRTSGRPVWLGSVKSNIGHTQAAAGVAGVMK 2905

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db       2906 MVLALGRGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPWSRG-GRVRRAGVSSFGISGTN 2964

Qy        421 AHVILEEAPA 430
Db       2965 AHVIVEEAPA 2974
```

RESULT 3
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33

Query Match 71.9%; Score 1628.5; DB 9; Length 3739;
Best Local Similarity 72.6%; Pred. No. 2.5e-119;
Matches 312; Conservative 42; Mismatches 75; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959
```

```
Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 4
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             71.9%;  Score 1628.5;  DB 9;  Length 3739;
  Best Local Similarity   72.6%;  Pred. No. 2.5e-119;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| ||| |||||:: |||||||  |  |
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               ||||  |   : |  :  :|:  || | |::|||:||| | ||||||:||||||||||
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :|||||||||:|  ||||||||||| ||  |||:|:||| |  |||
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||| |||||||||||||||||||||||||||||||||:|||:| |:
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||||  |  | || |||:|||:|||||||||:|:||
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:|||| ||  :||||| |:|||| : ||   || |||| |||||:|||
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 5
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             71.9%;  Score 1628.5;  DB 9;  Length 3739;
  Best Local Similarity   72.6%;  Pred. No. 2.5e-119;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| ||| |||||:: |||||||  |  |
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               ||||  |   : |  :  :|:  || | |::|||:||| | ||||||:||||||||||
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :|||||||||:|  ||||||||||| ||  |||:|:||| |  |||
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||| |||||||||||||||||||||||||||||||||:|||:| |:
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||||  |  | || |||:|||:|||||||||:|:||
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:|||| ||  :||||| |:|||| : ||   || |||| |||||:|||
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 6
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
```

```
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-2

  Query Match             71.9%;  Score 1628.5;  DB 9;  Length 3739;
  Best Local Similarity   72.6%;  Pred. No. 2.5e-119;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| ||| ||||:: |||||||  | |
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||  |   : |  :  :|:  || | |::|||:||| | |||||:|||||||||||
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :||||||||||:|  |||||||||| ||  |||:|:||| |  |||
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||| ||||||||||||||||||||||||||||||||:|||:| |:
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||| |  | || |||:|||:|||||||||:|:||
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:|||| ||  :||||| |:|||| : ||   || |||| |||||:|||
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 7
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             71.9%;  Score 1628.5;  DB 10;  Length 3739;
  Best Local Similarity   72.6%;  Pred. No. 2.5e-119;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| ||| ||||:: |||||||  | |
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||  |   : |  :  :|:  || | |::|||:||| | |||||:|||||||||||
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :||||||||||:|  |||||||||| ||  |||:|:||| |  |||
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||| ||||||||||||||||||||||||||||||||:|||:| |:
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||| |  | || |||:|||:|||||||||:|:||
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:|||| ||  :||||| |:|||| : ||   || |||| |||||:|||
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

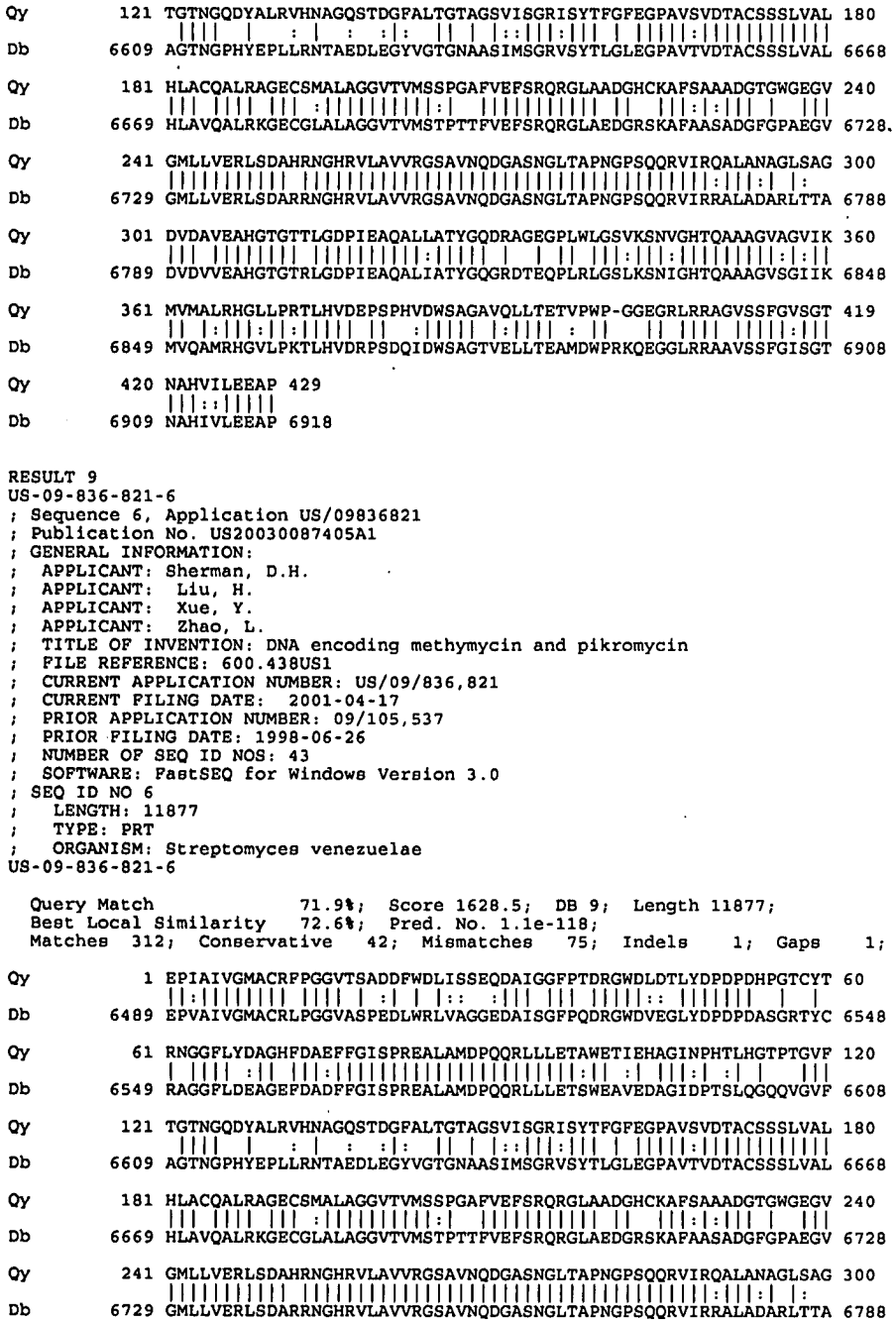
Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 8
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match             71.9%;  Score 1628.5;  DB 9;  Length 11877;
  Best Local Similarity   72.6%;  Pred. No. 1.1e-118;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

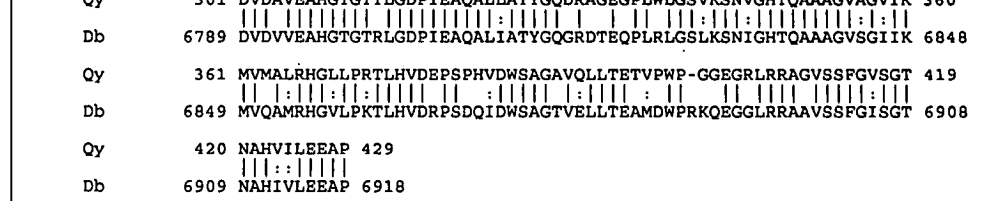
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| ||| ||||:: |||||||  | |
Db       6489 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 6548

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       6549 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 6608
```



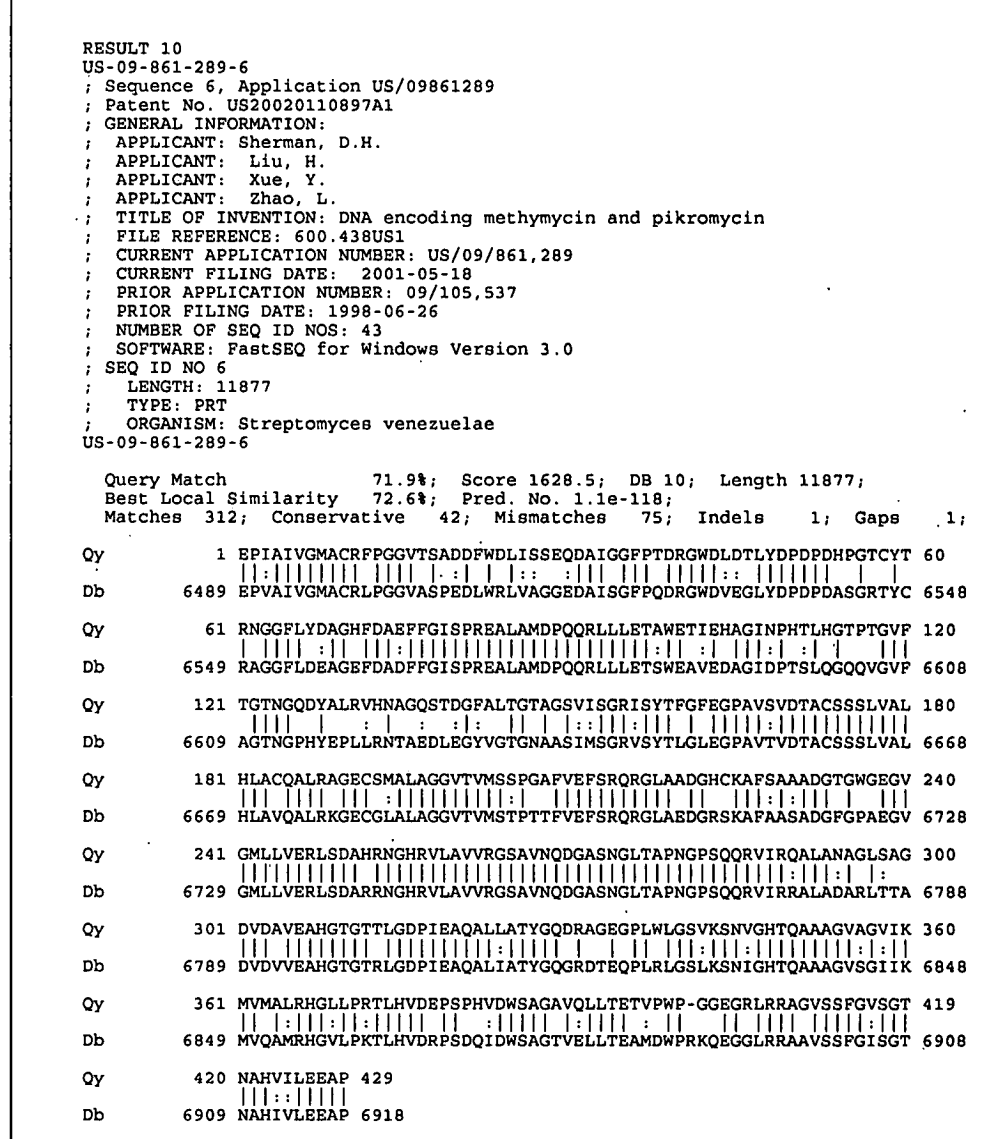
RESULT 9
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-6

Query Match 71.9%; Score 1628.5; DB 9; Length 11877;
Best Local Similarity 72.6%; Pred. No. 1.1e-118;
Matches 312; Conservative 42; Mismatches 75; Indels 1; Gaps 1;



RESULT 10
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6

Query Match 71.9%; Score 1628.5; DB 10; Length 11877;
Best Local Similarity 72.6%; Pred. No. 1.1e-118;
Matches 312; Conservative 42; Mismatches 75; Indels 1; Gaps 1;



```
RESULT 11
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             71.9%;  Score 1628.5;  DB 9;  Length 12199;
  Best Local Similarity   72.6%;  Pred. No. 1.1e-118;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| ||| |||||:: |||||||  |  |
Db       6811 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 6870

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||:|| :| |||:| :|  |   |||
Db       6871 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 6930

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||  |   : |  :  :|:  || | |::|||:||| | |||||:||||||||||||
Db       6931 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 6990

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :|||||||||:|  ||||||||||| ||  |||:|:||| |  |||
Db       6991 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 7050

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||||| |||||||||||||||||||||||||||||||||:|||:| |:
Db       7051 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 7110

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||||  |  | || |||:|||:|||||||||:|:||
Db       7111 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 7170

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:||||| ||  :||||| |:|||| : ||   || |||| |||||:|||
Db       7171 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 7230

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       7231 NAHIVLEEAP 7240


RESULT 12
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-1

  Query Match             71.8%;  Score 1625.5;  DB 9;  Length 4551;
  Best Local Similarity   72.3%;  Pred. No. 5.4e-119;
  Matches  311;  Conservative   41;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::  :||| || |||||:: || |||:|||| |
Db       1047 EPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYV 1106

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||: :   ||| ||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1107 RQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVF 1166

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    :|   : : |:  ||: |||   ||:|||:||| | ||||::||||||||||||
Db       1167 TGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVAL 1226

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||  ||||||| || :|| |||||||||| ||. |||:|:|||| | |||
Db       1227 HLAVQALRKGEVDMALAGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGV 1286

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||:|||||||||||||||||||||||||||||:|||:| |:
Db       1287 GVLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTS 1346

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:|||||  |  | || |||:|||:|||||||||:||||
Db       1347 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIK 1406

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:||||||:||||||||  :|||||||:|||| | ||   :| |||| |||||:|||
Db       1407 MVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGT 1466

Qy        420 NAHVILEEAP 429
              ||||:|||||
Db       1467 NAHVVLEEAP 1476


RESULT 13
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
```

```
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-31

  Query Match             71.6%;  Score 1622.5;  DB 9;  Length 4613;
  Best Local Similarity   72.1%;  Pred. No. 9.5e-119;
  Matches  310;  Conservative   42;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::   |||  || |||||:: || |||:|||| |
Db       1109 EPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYV 1168

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||: :   ||| ||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1169 RQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVF 1228

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    :|   : : |:  ||: |||   ||:|||:||| | ||||::||||||||||
Db       1229 TGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVAL 1288

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||  |||||| || :|| |||||||||| ||  |||:|:|||| | |||
Db       1289 HLAVQALRKGEVDMALAGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGV 1348

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||:|||||||:|||||||||||||||||||||:|||:| |:
Db       1349 GVLLVERLSDARRNGHQVLAVVRGSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTS 1408

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| |||||||||:||||| |  | || |||:|||:|||||||||:||||
Db       1409 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIK 1468

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:||||||:|||||||  :|||||||:|||| | ||   :| |||| ||||:|||
Db       1469 MVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGT 1528

Qy        420 NAHVILEEAP 429
              ||||:|||||
Db       1529 NAHVVLEEAP 1538


RESULT 14
US-09-988-384B-31
; Sequence 31, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-31

  Query Match             71.6%;  Score 1622.5;  DB 9;  Length 4613;
  Best Local Similarity   72.1%;  Pred. No. 9.5e-119;
  Matches  310;  Conservative   42;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::   |||  || |||||:: || |||:|||| |
Db       1109 EPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYV 1168

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||: :   ||| ||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1169 RQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVF 1228

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    :|   : : |:  ||: |||   ||:|||:||| | ||||::||||||||||
Db       1229 TGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVAL 1288

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||  |||||| || :|| |||||||||| ||  |||:|:|||| | |||
Db       1289 HLAVQALRKGEVDMALAGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGV 1348

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||:|||||||:|||||||||||||||||||||:|||:| |:
Db       1349 GVLLVERLSDARRNGHQVLAVVRGSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTS 1408

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| |||||||||:||||| |  | || |||:|||:|||||||||:||||
Db       1409 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIK 1468

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:||||||:|||||||  :|||||||:|||| | ||   :| |||| ||||:|||
Db       1469 MVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGT 1528

Qy        420 NAHVILEEAP 429
              ||||:|||||
Db       1529 NAHVVLEEAP 1538


RESULT 15
US-09-836-821-31
; Sequence 31, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-31

  Query Match             71.6%;  Score 1622.5;  DB 9;  Length 4613;
  Best Local Similarity   72.1%;  Pred. No. 9.5e-119;
  Matches  310;  Conservative   42;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:|||||||| |||| | :| | |::   |||  || |||||:: || |||:|||| |
Db       1109 EPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYV 1168

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||: :   ||| ||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1169 RQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVF 1228

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    :|   : : |:  ||: |||   ||:|||:||| | ||||::||||||||||
Db       1229 TGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVAL 1288
```

```
Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||  ||||||| || :|| |||||||||||| ||  |||:|:|||| | |||
Db       1289 HLAVQALRKGEVDMALAGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGV 1348

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| |||:|||||||||:|||||||||||||||||||||:|||:| |:
Db       1349 GVLLVERLSDARRNGHQVLAVVRGSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTS 1408

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:||||| |  | || |||:|||:||||||||:||||
Db       1409 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIK 1468

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:||||||:||||||||  :|||||||:|||| | ||   :| |||| |||||:|||
Db       1469 MVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGT 1528

Qy        420 NAHVILEEAP 429
              ||||:|||||
Db       1529 NAHVVLEEAP 1538
```

Search completed: June 17, 2003, 13:46:45
Job time : 15.398 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 7.81547 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_2086_2515
Perfect score: 2265
Sequence: 1 EPIAIVGMACRFPGGVTSAD.....VSSFGVSGTNAHVILEEAPA 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1704	75.2	3170	4	US-09-036-987A-4	Sequence 4, Appli
2	1704	75.2	3170	4	US-09-370-700-4	Sequence 4, Appli
3	1672	73.8	4928	4	US-09-036-987A-5	Sequence 5, Appli
4	1672	73.8	4928	4	US-09-370-700-5	Sequence 5, Appli
5	1654	73.0	3519	4	US-09-428-517-4	Sequence 4, Appli
6	1646.5	72.7	4150	4	US-09-428-517-2	Sequence 2, Appli
7	1640	72.4	1996	2	US-08-804-227C-9	Sequence 9, Appli
8	1640	72.4	1996	2	US-08-804-198-3	Sequence 3, Appli
9	1640	72.4	5588	4	US-09-036-987A-6	Sequence 6, Appli
10	1640	72.4	5588	4	US-09-370-700-6	Sequence 6, Appli
11	1631.5	72.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
12	1628.5	71.9	3739	3	US-09-320-878-2	Sequence 2, Appli
13	1628.5	71.9	3739	4	US-09-105-537-33	Sequence 33, Appl
14	1628.5	71.9	11877	4	US-09-105-537-6	Sequence 6, Appli
15	1625.5	71.8	4551	3	US-09-320-878-1	Sequence 1, Appli
16	1624.5	71.7	4545	2	US-08-804-227C-14	Sequence 14, Appl
17	1624.5	71.7	4550	2	US-08-804-227C-8	Sequence 8, Appli
18	1624.5	71.7	4550	2	US-08-804-198-2	Sequence 2, Appli
19	1622.5	71.6	2595	4	US-09-036-987A-2	Sequence 2, Appli
20	1622.5	71.6	2595	4	US-09-370-700-2	Sequence 2, Appli
21	1622.5	71.6	4613	4	US-09-105-537-31	Sequence 31, Appl
22	1619.5	71.5	2152	4	US-09-036-987A-3	Sequence 3, Appli
23	1619.5	71.5	2152	4	US-09-370-700-3	Sequence 3, Appli
24	1609	71.0	3816	4	US-09-428-517-3	Sequence 3, Appli
25	1597	70.5	3729	2	US-08-804-227C-4	Sequence 4, Appli
26	1587.5	70.1	3724	2	US-08-804-227C-10	Sequence 10, Appl
27	1587.5	70.1	3724	2	US-08-804-198-4	Sequence 4, Appli
28	1576.5	69.6	1891	2	US-08-804-227C-12	Sequence 12, Appl
29	1576.5	69.6	1891	2	US-08-804-198-6	Sequence 6, Appli
30	1574.5	69.5	1841	2	US-08-804-227C-6	Sequence 6, Appli
31	1570	69.3	5215	4	US-09-105-537-2	Sequence 2, Appli
32	1561.5	68.9	1562	3	US-09-320-878-3	Sequence 3, Appli
33	1561.5	68.9	1562	4	US-09-105-537-35	Sequence 35, Appl
34	1544.5	68.2	3567	2	US-07-642-734C-4	Sequence 4, Appli
35	1544.5	68.2	3567	3	US-08-439-009A-4	Sequence 4, Appli
36	1543	68.1	1346	3	US-09-320-878-4	Sequence 4, Appli
37	1538	67.9	1346	4	US-09-105-537-37	Sequence 37, Appl
38	1535	67.8	1864	2	US-08-804-227C-3	Sequence 3, Appli
39	1515	66.9	1611	2	US-08-804-227C-5	Sequence 5, Appli
40	1495.5	66.0	3491	2	US-07-642-734C-2	Sequence 2, Appli
41	1495.5	66.0	3491	3	US-08-439-009A-2	Sequence 2, Appli
42	1448	63.9	1580	2	US-08-804-227C-11	Sequence 11, Appl
43	1448	63.9	1580	2	US-08-804-198-5	Sequence 5, Appli
44	1431.5	63.2	3170	2	US-07-642-734C-5	Sequence 5, Appli
45	1431.5	63.2	3170	3	US-08-439-009A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-036-987A-4
; Sequence 4, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3170 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-4

Query Match 75.2%; Score 1704; DB 4; Length 3170;
Best Local Similarity 75.3%; Pred. No. 3.5e-147;
Matches 323; Conservative 38; Mismatches 66; Indels 2; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| | ||:||||    :  | |::|  |||| || |||| || ||||||| ||||||
Db         34 EPIAIVAMGCRYPGGVQDPEGLWKLVASGGDAIGEFPADRGWHLDELYDPDPDQPGTCYT 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| ||| || |||||||||||||||||| :|||:| ||::| :| |: ||||
Db         94 RHGGFLHDAGEFDAGFFDISPREALAMDPQQRLLLEISWETVESAGMDPRSLRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |   : |    | ||:  :|:  || |||| |||::| |||||||:||||||||||||
Db        154 AGLMYEGYDTGAHRAGEGVEGYLGTGNAGSVASGRVAYAFGFEGPAVTVDTACSSSLVAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:|| ||| :|||||||||:|  |||||||||| || ||:|:|||||||||||
Db        214 HLACQSLRQGECDLALAGGVTVMSTPERFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGA 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |::|:|||||| ||||||||||||||||||||||||||| :|:|||:| | :|||||
Db        274 GLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGLAQERVIQQVLTSAGLSAS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| ||||||||||:| |||||  : ||||||||||:|||||||||||||||
Db        334 DVDAVEAHGTGTRLGDPIEAQALIAAYGQDRDRDRPLWLGSVKSNIGHTQAAAGVAGVIK 393

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:||| |||||||||: |||||||||:|||| : |||      ||||||||||||||
Db        394 MVMAMRHGELPRTLHVDEPNSHVDWSAGAVRLLTENIRWPG--TGTRRAGVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              |||||| |
Db        452 AHVILEHDP 460
```

RESULT 2
US-09-370-700-4
; Sequence 4, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3170
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-4

Query Match 75.2%; Score 1704; DB 4; Length 3170;
Best Local Similarity 75.3%; Pred. No. 3.5e-147;
Matches 323; Conservative 38; Mismatches 66; Indels 2; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| | ||:||||    :  | |::|  |||| || |||| || ||||||| ||||||
Db         34 EPIAIVAMGCRYPGGVQDPEGLWKLVASGGDAIGEFPADRGWHLDELYDPDPDQPGTCYT 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| ||| || |||||||||||||||||| :|||:| ||::| :| |: ||||
Db         94 RHGGFLHDAGEFDAGFFDISPREALAMDPQQRLLLEISWETVESAGMDPRSLRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |   : |    | ||:  :|:  || |||| |||::| |||||||:||||||||||||
Db        154 AGLMYEGYDTGAHRAGEGVEGYLGTGNAGSVASGRVAYAFGFEGPAVTVDTACSSSLVAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:|| ||| :|||||||||:|  |||||||||| || ||:|:|||||||||||
Db        214 HLACQSLRQGECDLALAGGVTVMSTPERFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGA 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |::|:|||||| ||||||||||||||||||||||||||| :|:|||:| | :|||||
Db        274 GLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGLAQERVIQQVLTSAGLSAS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| ||||||||||:| |||||  : ||||||||||:|||||||||||||||
Db        334 DVDAVEAHGTGTRLGDPIEAQALIAAYGQDRDRDRPLWLGSVKSNIGHTQAAAGVAGVIK 393

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:||| |||||||||: |||||||||:|||| : |||      ||||||||||||||
Db        394 MVMAMRHGELPRTLHVDEPNSHVDWSAGAVRLLTENIRWPG--TGTRRAGVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              |||||| |
Db        452 AHVILEHDP 460
```

RESULT 3
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4928 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-5

Query Match 73.8%; Score 1672; DB 4; Length 4928;
Best Local Similarity 75.2%; Pred. No. 5.8e-144;

```
  Matches  322;  Conservative   34;  Mismatches   70;  Indels    2;  Gaps    2;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||||| |||||||:| :: | |::   ||: ||| ||||||  |||||||  || |
Db       1787 VAIVGMGCRFPGGVSSPEELWRLVAGGVDAVAGFPDDRGWDLAALYDPDPDRLGTSYVCE 1846

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTG 122
              |||| ||  |||: |||||||||||||||||||| ||||:| |||:| :|||: |||| |
Db       1847 GGFLRDAAEFDADMFGISPREALAMDPQQRLLLEVAWETLERAGIDPFSLHGSRTGVFAG 1906

Qy        123 TNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
                 ||  | :  | :  :|   || ||||:|||::|:||||||||:|||||||||||||
Db       1907 LMYHDYGARFITRAPEGFEGHLGTGNAGSVLSGRVAYSFGFEGPAVTVDTACSSSLVALH 1966

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              || |||||||  ||||||||||:|  ||||||||||| || ||:|:|||||||||||| |
Db       1967 LAGQALRAGECEFALAGGVTVMSTPTTFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGAG 2026

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
              ::|:|||||| |||| ||||||||||||||||||||||||||||| ||| :||||  |
Db       2027 LVLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALTSAGLSVSD 2086

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              |||||||||||| ||||||||||:||||:||    ||||||||||:|||||||||||||
Db       2087 VDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAGVIKM 2146

Qy        362 VMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNA 421
              |||:| | ||||||||||  |||||| ||||||  |||   ||||||||||||:|||||
Db       2147 VMAMRQGELPRTLHVDEPSAQVDWSAGTVQLLTENTPWP-DSGRLRRAGVSSFGISGTNA 2205

Qy        422 HVILEEAP 429
              |:|||: |
Db       2206 HLILEQPP 2213


RESULT 4
US-09-370-700-5
; Sequence 5, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 4928
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-5

  Query Match             73.8%;  Score 1672;  DB 4;  Length 4928;
  Best Local Similarity   75.2%;  Pred. No. 5.8e-144;
  Matches  322;  Conservative   34;  Mismatches   70;  Indels    2;  Gaps    2;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||||| |||||||:| :: | |::   ||: ||| ||||||  |||||||  || |
Db       1787 VAIVGMGCRFPGGVSSPEELWRLVAGGVDAVAGFPDDRGWDLAALYDPDPDRLGTSYVCE 1846

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTG 122
              |||| ||  |||: |||||||||||||||||||| ||||:| |||:| :|||: |||| |
Db       1847 GGFLRDAAEFDADMFGISPREALAMDPQQRLLLEVAWETLERAGIDPFSLHGSRTGVFAG 1906

Qy        123 TNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
                 ||  | :  | :  :|   || ||||:|||::|:||||||||:|||||||||||||
Db       1907 LMYHDYGARFITRAPEGFEGHLGTGNAGSVLSGRVAYSFGFEGPAVTVDTACSSSLVALH 1966

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              || |||||||  ||||||||||:|  ||||||||||| || ||:|:|||||||||||| |
Db       1967 LAGQALRAGECEFALAGGVTVMSTPTTFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGAG 2026

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
              ::|:|||||| |||| ||||||||||||||||||||||||||||| ||| :||||  |
Db       2027 LVLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALTSAGLSVSD 2086

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              |||||||||||| ||||||||||:||||:||    ||||||||||:|||||||||||||
Db       2087 VDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAGVIKM 2146

Qy        362 VMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNA 421
              |||:| | ||||||||||  |||||| ||||||  |||   ||||||||||||:|||||
Db       2147 VMAMRQGELPRTLHVDEPSAQVDWSAGTVQLLTENTPWP-DSGRLRRAGVSSFGISGTNA 2205

Qy        422 HVILEEAP 429
              |:|||: |
Db       2206 HLILEQPP 2213


RESULT 5
US-09-428-517-4
; Sequence 4, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 3519
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-4

  Query Match             73.0%;  Score 1654;  DB 4;  Length 3519;
  Best Local Similarity   73.3%;  Pred. No. 1.6e-142;
  Matches  315;  Conservative   40;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              | |||  |:||| ||: | :| |  ::  :||: | | ||||||| || |||::||| |
Db       1687 EAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGTTYV 1746

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | | |||  ||| |||||||||||||||||||||:||  | | |:|:|: || ||:|
Db       1747 REGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTATGIF 1806

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |   | |       | :|  |: ||||| :|:|||||||| |||||:||||||||||
Db       1807 IGAGHQGYGPDPKRAPESVAGYLLTGTASAVLSGRISYTFGLEGPAVTVDTACSSSLVAL 1866

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| |||:|:|||| |||:| |||||||:|:| || |||||||||| ||||||
Db       1867 HLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAADGMGWGEGV 1926
```

```
Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       1927 SLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 1986

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       1987 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAAAGVAGVMK 2046

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db       2047 MVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGISAFGVSGTN 2106

Qy        421 AHVILEEAPA 430
Db       2107 AHVIVEEAPA 2116
```

RESULT 6
US-09-428-517-2
; Sequence 2, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-2

Query Match 72.7%; Score 1646.5; DB 4; Length 4150;
Best Local Similarity 72.8%; Pred. No. 9.9e-142;
Matches 313; Conservative 41; Mismatches 75; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db       2546 EPIAIVGMACRFPGGIASPEDLWDVLTAGEDVVSGLPQNRGWDLGRLYDPDPDRAGTSYM 2605

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db       2606 REGAFLHEAGEFDAAFFGISPREALAMDPQQRLLLETSWEALERAGITPSKLAGSPTGVF 2665

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       2666 FGMSNQDYAAQAGDVPSELEGYLLTGSISSVASGRVAYTFGLEGPAVTVDTACSSSLVAL 2725

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       2726 HLAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCKAFAASADGFGAAEGV 2785

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       2786 GVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 2845

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       2846 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRTSGRPVWLGSVKSNIGHTQAAAGVAGVMK 2905

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db       2906 MVLALGRGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPWSRG-GRVRRAGVSSFGISGTN 2964

Qy        421 AHVILEEAPA 430
Db       2965 AHVIVEEAPA 2974
```

RESULT 7
US-08-804-227C-9
; Sequence 9, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1996 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-9

Query Match 72.4%; Score 1640; DB 2; Length 1996;
Best Local Similarity 71.9%; Pred. No. 1.4e-141;
Matches 309; Conservative 46; Mismatches 75; Indels 0; Gaps 0;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db         35 EPIAIVGMACRYPGGVASPDDLWDLVAAGTDAVSAFPVDRGWDVEGLYDPDPEAVGRSYV 94

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db         95 REGGFLHSAAEFDAEFFGISPREAAAMDPQQRLLLETSWEALERAGIVPASLRGTRTGVF 154

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db        155 TGVMYDDYGSRFDSAPPEYEGYLVNGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL 214

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db        215 HLAVQSLRRGECDLALAGGVTVMATPTVLVEFSRQRGLAADGRCKAFAEGADGTAWAEGV 274

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIREALADAGLTPA 334
Db        275 GVLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPSGPAQQRVIREALADAGLTPA 334
```

```
Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| ||||||| ||||||| :| |:||||||:|||:|| |||||| ||||
Db        335 DVDAVEAHGTGTPLGDPIEAGALLATYGSERQGQGPLWLGSLKSNIGHAQAAAGVGGVIK 394

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              :| |:||| |||||||| ||  |:|::|||:|||||  ||     |:||| ||:|||||||
Db        395 VVQAMRHGSLPRTLHVDAPSSKVEWASGAVELLTETRSWPRRVERVRRAAVSAFGVSGTN 454

Qy        421 AHVILEEAPA 430
              |||:||||||
Db        455 AHVVLEEAPA 464


RESULT 8
US-08-804-198-3
; Sequence 3, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1996 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-3

  Query Match           72.4%;  Score 1640;  DB 2;  Length 1996;
  Best Local Similarity  71.9%;  Pred. No. 1.4e-141;
  Matches  309;  Conservative   46;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||:|||| | || ||||:::  ||:  || |||||:: ||||||:  |  |
Db         35 EPIAIVGMACRYPGGVASPDDLWDLVAAGTDAVSAFPVDRGWDVEGLYDPDPEAVGRSYV 94

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | ||||: |  |||||||||||||| ||||||||||||:|| :| ||| | :| || ||||
Db         95 REGGFLHSAAEFDAEFFGISPREAAAMDPQQRLLLETSWEALERAGIVPASLRGTRTGVF 154

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    ||  |  :|    :|: : |:|||: |||::|  | ||||::||||||||||||
Db        155 TGVMYDDYGSRFDSAPPEYEGYLVNGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL 214

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |:|| ||| :||||||||::|  ||||||||||||| ||||:  |||| | |||
Db        215 HLAVQSLRRGECDLALAGGVTVMATPTVLVEFSRQRGLAADGRCKAFAEGADGTAWAEGV 274

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| ||||||||||||||||||||||||||||:||:|||||:|||:|||:
Db        275 GVLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPSGPAQQRVIREALADAGLTPA 334

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| ||||||| ||||||| :| |:||||||:|||:|| |||||| ||||
Db        335 DVDAVEAHGTGTPLGDPIEAGALLATYGSERQGQGPLWLGSLKSNIGHAQAAAGVGGVIK 394

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              :| |:||| |||||||| ||  |:|::|||:|||||  ||     |:||| ||:|||||||
Db        395 VVQAMRHGSLPRTLHVDAPSSKVEWASGAVELLTETRSWPRRVERVRRAAVSAFGVSGTN 454

Qy        421 AHVILEEAPA 430
              |||:||||||
Db        455 AHVVLEEAPA 464


RESULT 9
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match           72.4%;  Score 1640;  DB 4;  Length 5588;
  Best Local Similarity  73.2%;  Pred. No. 6e-141;
  Matches  314;  Conservative   40;  Mismatches   73;  Indels    2;  Gaps    2;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              :|: |||| ||:||||:| :: | |::   ||:. || |||||   |:|||||  || |
```

```
Db       1782 DPVVIVGMGCRYPGGVSSPEELWRLVAGGLDAVAEFPDDRGWDQAGLFDPDPDRLGTSYV 1841

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120

Db       1842 CEGGFLRDAAEFDAGFFGISPREALAMDPQQRLLLEVAWETVERAGIDPLSLRGSRTGVF 1901

Qy        121 TGTNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179

Db       1902 AGLMHHDYGARFITRAPEGFEGYLGNGSAGGVFSGRVAYSFGFEGPAVTVDTACSSSLVA 1961

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239

Db       1962 LHLAGQALRSGECDLALAGGVTVMATPGMFVEFSRQRGLAADGRCKSFAAAADGTGWGEG 2021

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299

Db       2022 AGLVLLERLSDARRNGHAVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALASAGLSV 2081

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 359

Db       2082 SDVDAVEAHGTGTRLGDPIEAQALIATYGQGRDSDRPLWLGSVKSNIGHTQAAAGVAGVI 2141

Qy        360 KMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 419

Db       2142 KMVMAMRHGQLPATLHVDEPTSEVDWSAGDVQLLTENTPWP-GNSHPRRVGVSSFGISGT 2200

Qy        420 NAHVILEEA 428

Db       2201 NAHVILEQA 2209


RESULT 10
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match             72.4%;  Score 1640;  DB 4;  Length 5588;
  Best Local Similarity   73.2%;  Pred. No. 6e-141;
  Matches  314;  Conservative   40;  Mismatches   73;  Indels    2;  Gaps    2;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60

Db       1782 DPVVIVGMGCRYPGGVSSPEELWRLVAGGLDAVAEFPDDRGWDQAGLFDPDPDRLGTSYV 1841

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120

Db       1842 CEGGFLRDAAEFDAGFFGISPREALAMDPQQRLLLEVAWETVERAGIDPLSLRGSRTGVF 1901

Qy        121 TGTNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179

Db       1902 AGLMHHDYGARFITRAPEGFEGYLGNGSAGGVFSGRVAYSFGFEGPAVTVDTACSSSLVA 1961

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239

Db       1962 LHLAGQALRSGECDLALAGGVTVMATPGMFVEFSRQRGLAADGRCKSFAAAADGTGWGEG 2021

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299

Db       2022 AGLVLLERLSDARRNGHAVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALASAGLSV 2081

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 359

Db       2082 SDVDAVEAHGTGTRLGDPIEAQALIATYGQGRDSDRPLWLGSVKSNIGHTQAAAGVAGVI 2141

Qy        360 KMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 419

Db       2142 KMVMAMRHGQLPATLHVDEPTSEVDWSAGDVQLLTENTPWP-GNSHPRRVGVSSFGISGT 2200

Qy        420 NAHVILEEA 428

Db       2201 NAHVILEQA 2209


RESULT 11
US-08-804-227C-2
; Sequence 2, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
;   APPLICANT:  DeHoff, Bradley S.
;   APPLICANT:  Kuhstoss, Stuart A.
;   APPLICANT:  Rosteck, Paul R., Jr.
;   APPLICANT:  Sutton, Kimberly L.
;   TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;   NUMBER OF SEQUENCES:  15
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  THOMAS G. PLANT 1501
;     STREET:  LILLY CORPORATE CENTER
;     CITY:  INDIANAPOLIS
;     STATE:  IN
;     COUNTRY:  USA
;     ZIP:  46285
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  MS-DOS
;     SOFTWARE:  ASCI(DOS) Text only
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/804,227C
;     FILING DATE:  February 21, 1997
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Plant, Thomas, G.
;     REGISTRATION NUMBER:  35,784
;     REFERENCE/DOCKET NUMBER:  X-8231
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  4472 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  unknown
;   MOLECULE TYPE:  protein
US-08-804-227C-2

  Query Match             72.0%;  Score 1631.5;  DB 2;  Length 4472;
  Best Local Similarity   71.9%;  Pred. No. 2.6e-140;
  Matches  309;  Conservative   46;  Mismatches   74;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60

Db       2576 DPVVIVGMACRFPGGAGTPEALWKLVTEERDVIGAAPTDRGWDLDSVYDPEPGVAGKTYV 2635

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120

Db       2636 REGGFLHDAAEFDAEFFGISPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVY 2695
```

```
Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       2696 VGLTHQEYASRLHEAPEEYEGYLLTGKSASVVSGRISYTLGLEGPSLSIDTACSSSLVAL 2755

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       2756 HNAAQALRGGECDMALAGGVTVMAAPGLFVEFSRQRGLAADGRCKAFADGADGTAWAEGA 2815

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       2816 GVVLVERLSDARRLGHPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALANARLTVA 2875

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       2876 DVDVVEAHGTGTRLGDPIEAQALLGTYGRDRDAECPVWLGSLKSNIGHAQAAAGVAGVIK 2935

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
Db       2936 MVLAMRYGWLPRTLHVDEPSRHVDWSAGGVRLLTEAREWPGVD-RPRRAAVSAFGVSGTN 2994

Qy        421 AHVILEEAPA 430
Db       2995 AHLILEAPEA 3004
```

RESULT 12
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2

Query Match 71.9%; Score 1628.5; DB 3; Length 3739;
Best Local Similarity 72.6%; Pred. No. 3.8e-140;
Matches 312; Conservative 42; Mismatches 75; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

Qy        420 NAHVILEEAP 429
Db       1960 NAHIVLEEAP 1969
```

RESULT 13
US-09-105-537-33
; Sequence 33, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-33

Query Match 71.9%; Score 1628.5; DB 4; Length 3739;
Best Local Similarity 72.6%; Pred. No. 3.8e-140;
Matches 312; Conservative 42; Mismatches 75; Indels 1; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
Db       1540 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 1599

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
Db       1600 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 1659

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
Db       1660 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 1719

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
Db       1720 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 1779

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
Db       1780 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 1839

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
Db       1840 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 1899
```

```
Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:||||| ||  :||||| |:|||| : ||    || |||| |||||:|||
Db       1900 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 1959

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       1960 NAHIVLEEAP 1969


RESULT 14
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             71.9%;  Score 1628.5;  DB 4;  Length 11877;
  Best Local Similarity   72.6%;  Pred. No. 2e-139;
  Matches  312;  Conservative   42;  Mismatches   75;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||||||||| |||| | :| | |::  :||| ||| |||||:: |||||||   |  |
Db       6489 EPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYC 6548

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||| :|| |||:||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       6549 RAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVF 6608

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||  |   : |  :  :|:  || | |::|||:||| | |||||:||||||||||||
Db       6609 AGTNGPHYEPLLRNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVAL 6668

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||| :||||||||||:|  ||||||||||| ||  |||:|:||| |  |||
Db       6669 HLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGV 6728

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||||| ||||||||||||||||||||||||||||||||||:|||:| |:
Db       6729 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTA 6788

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:||||| |  | || |||:|||:|||||||||:|:||
Db       6789 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIK 6848

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:|||:||:||||| ||  :||||| |:|||| : ||    || |||| |||||:|||
Db       6849 MVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGT 6908

Qy        420 NAHVILEEAP 429
              |||::|||||
Db       6909 NAHIVLEEAP 6918


RESULT 15
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-1

  Query Match             71.8%;  Score 1625.5;  DB 3;  Length 4551;
  Best Local Similarity   72.3%;  Pred. No. 9.6e-140;
  Matches  311;  Conservative   41;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||||||||| |||| | :| | |::   |||  || |||||:: || |||:|||| |
Db       1047 EPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYV 1106

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | |||: :   ||| ||||||||||||||||||||||:|| :| |||:| :| |   |||
Db       1107 RQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVF 1166

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||    :|   : : |:  ||: |||   ||:|||:||| | ||||::||||||||||||
Db       1167 TGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVAL 1226

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||  ||||||| || :|| |||||||||| ||  |||:|:|||| | |||
Db       1227 HLAVQALRKGEVDMALAGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGV 1286

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||| |||||||||||||||||||||||||||||||||||:|||:| |:
Db       1287 GVLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTS 1346

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| ||||||||||:||||| |  | || |||:|||:|||||||||:||||
Db       1347 DVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIK 1406

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGT 419
              || |:||||||:||||||  :|||||||:|||| | ||   :| |||| |||||:|||
Db       1407 MVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGT 1466

Qy        420 NAHVILEEAP 429
              ||||:|||||
Db       1467 NAHVVLEEAP 1476


Search completed: June 17, 2003, 13:09:29
Job time : 8.81548 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 21.0646 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_2086_2515
Perfect score: 2265
Sequence: 1 EPIAIVGMACRFPGGVTSAD.....VSSFGVSGTNAHVILEEAPA 430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2265	100.0	3972	21	AAB23749	S. avermitilis ave
2	2265	100.0	3972	22	AAG65264	Streptomyces averm
3	2265	100.0	3972	22	AAG65268	Streptomyces averm
4	2042	90.2	6239	21	AAB23750	S. avermitilis ave
5	2042	90.2	6239	22	AAG65265	Streptomyces averm
6	1968.5	86.9	5532	21	AAB23752	S. avermitilis ave
7	1968.5	86.9	5532	22	AAG65267	Streptomyces averm
8	1744.5	77.0	4881	21	AAB23751	S. avermitilis ave
9	1744.5	77.0	4881	22	AAG65266	Streptomyces averm
10	1704	75.2	3170	20	AAY39299	SpnC a polyketide
11	1704	75.2	3170	22	AAB70967	S. spinosa protein
12	1672	73.8	4924	22	AAB70968	S. spinosa protein
13	1672	73.8	4928	20	AAY39300	SpnD a polyketide
14	1656	73.1	11096	22	AAE10129	Streptomyces nours
15	1654	73.0	3519	21	AAY92709	S. antibioticus 8,
16	1646.5	72.7	4150	21	AAY92707	S. antibioticus 8,
17	1645.5	72.6	6797	22	AAB31558	Pimaricin biosynth
18	1640	72.4	1996	18	AAW22607	Platenolide syntha
19	1640	72.4	1996	18	AAW23717	Platenolide syntha
20	1640	72.4	5588	20	AAY39301	SpnE a polyketide
21	1640	72.4	5588	22	AAB70969	S. spinosa protein
22	1636	72.2	5069	19	AAW52846	A. mediterranei ri
23	1631.5	72.0	4472	18	AAW22601	Tylactone synthase
24	1628.5	71.9	3739	21	AAB18638	Amino acid sequenc
25	1628.5	71.9	3739	21	AAY77193	S. venezuelae macr
26	1628.5	71.9	3739	21	AAY77201	S. venezuelae pik
27	1628.5	71.9	3739	21	AAY67202	Narbonolide syntha
28	1628.5	71.9	12199	21	AAY77180	S. venezuelae pik
29	1625.5	71.8	4551	21	AAB18637	Amino acid sequenc
30	1625.5	71.8	4551	21	AAY67201	Narbonolide syntha
31	1624.5	71.7	4545	18	AAW22611	Hybrid srmG/tylG O
32	1624.5	71.7	4550	18	AAW23716	Platenolide syntha
33	1624.5	71.7	4550	18	AAW22606	Platenolide syntha
34	1622.5	71.6	2595	20	AAY39297	SpnA a polyketide
35	1622.5	71.6	2595	22	AAB70965	S. spinosa protein
36	1622.5	71.6	4613	21	AAY77192	S. venezuelae macr
37	1622.5	71.6	4613	21	AAY77200	S. venezuelae pik
38	1619.5	71.5	2152	20	AAY39298	SpnB a polyketide
39	1619.5	71.5	2152	22	AAB70966	S. spinosa protein
40	1614.5	71.3	7068	22	AAE10142	Streptomyces nours
41	1614.5	71.3	9477	22	AAE10144	Streptomyces nours
42	1611	71.1	4572	19	AAW52845	A. mediterranei ri
43	1609	71.0	3816	21	AAY92708	S. antibioticus 8,
44	1605	70.9	888	23	AAU76935	Hybrid polyketide
45	1597	70.5	3729	18	AAW22603	Tylactone synthase

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

```
PT   agrochemical use -
XX
PS   Claim 32; Page 203-223; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   3972 AA;

  Query Match          100.0%;  Score 2265;  DB 21;  Length 3972;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-192;
  Matches  430;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2086 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2145

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2146 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2205

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2206 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 2265

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2266 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 2325

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2326 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 2385

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2386 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 2445

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2446 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 2505

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       2506 AHVILEEAPA 2515


RESULT 2
AAG65264
ID   AAG65264 standard; Protein; 3972 AA.
XX
AC   AAG65264;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79277.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 167-180; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   3972 AA;

  Query Match          100.0%;  Score 2265;  DB 22;  Length 3972;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-192;
  Matches  430;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2086 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2145

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2146 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2205

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2206 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 2265

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2266 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 2325

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2326 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 2385

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2386 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 2445

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2446 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 2505

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       2506 AHVILEEAPA 2515


RESULT 3
AAG65268
ID   AAG65268 standard; Protein; 3972 AA.
XX
AC   AAG65268;
XX
```

DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 2265; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1.1e-192;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2086 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2145

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2146 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2205

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2206 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 2265

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2266 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 2325

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2326 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 2385

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2386 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 2445

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2446 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 2505

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       2506 AHVILEEAPA 2515
```

RESULT 4
AAB23750
ID AAB23750 standard; Protein; 6239 AA.
XX
AC AAB23750;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 32; Page 223-254; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 6239 AA;

Query Match 90.2%; Score 2042; DB 21; Length 6239;
Best Local Similarity 89.5%; Pred. No. 1.8e-172;
Matches 385; Conservative 14; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2680 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2739

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2740 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2799

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |||||||:|  :  |   |:|| ||| | |: ||||||  | ||||||::||||||||||
Db       2800 TGTNGQDHAAHIRQAPSGTEGFVLTGAATSIASGRISYILGLEGPAVTLDTACSSSLVAL 2859
```

```
Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:||:|||:|||||| |||::|  | ||:|||||| || ||||||||||||||||||
Db       2860 HLACQSLRSGECTMALAGGATVMTTPITFTEFARQRGLAPDGRCKAFSAAADGTGWGEGV 2919

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||| |:
Db       2920 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANADLTPA 2979

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||||||||||:|||||||| | ||||||||||||||||||||||||||||||
Db       2980 DVDAVEAHGTGTTLGDPIEAQAILATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIK 3039

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||| || ||| |||||||||||||||||||||||||||||| ||||||||||||||
Db       3040 MVMALRHRTLPPTLHADEPSPHVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTN 3099

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       3100 AHVILEEAPA 3109
```

RESULT 5
AAG65265
ID AAG65265 standard; Protein; 6239 AA.
XX
AC AAG65265;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 5.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 180-201; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 6239 AA;

 Query Match 90.2%; Score 2042; DB 22; Length 6239;
 Best Local Similarity 89.5%; Pred. No. 1.8e-172;
 Matches 385; Conservative 14; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2680 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2739

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2740 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2799

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |||||||:|  :  |   |:|| ||| | |: ||||||  | |||||::|||||||||||
Db       2800 TGTNGQDHAAHIRQAPSGTEGFVLTGAATSIASGRISYILGLEGPAVTLDTACSSSLVAL 2859

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:||:|||:|||||| |||::|  | ||:|||||| || ||||||||||||||||||
Db       2860 HLACQSLRSGECTMALAGGATVMTTPITFTEFARQRGLAPDGRCKAFSAAADGTGWGEGV 2919

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||| |:
Db       2920 GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANADLTPA 2979

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||||||||||:|||||||| | ||||||||||||||||||||||||||||||
Db       2980 DVDAVEAHGTGTTLGDPIEAQAILATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIK 3039

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||||| || ||| |||||||||||||||||||||||||||||| ||||||||||||||
Db       3040 MVMALRHRTLPPTLHADEPSPHVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTN 3099

Qy        421 AHVILEEAPA 430
              ||||||||||
Db       3100 AHVILEEAPA 3109
```

RESULT 6
AAB23752
ID AAB23752 standard; Protein; 5532 AA.
XX
AC AAB23752;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92302.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 32; Page 278-305; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;

```
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;

  Query Match             86.9%;  Score 1968.5;  DB 21;  Length 5532;
  Best Local Similarity   87.2%;  Pred. No. 5.7e-166;
  Matches  376;  Conservative   23;  Mismatches   31;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||:|||||:|||||||:: |:|::| :| :| ||||||||: |:||:|  ||| ||
Db       1864 EPVAIIGMACRYPGGVTSAEELWELLASGRDTVGEFPTDRGWDLEALFDPEPGRPGTSYT 1923

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | |||||| ||| ||||||||||||||||||||:|| :| |||:| |: |: ||||
Db       1924 RCGSFLYDAGEFDAGFFGISPREALAMDPQQRLLLEASWEAMEQAGIDPTTVRGSQTGVF 1983

Qy        121 TGTNGQDYALRVH-NAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179
               |   | |  |:| ||   |:|: ||||:||| ||||||||||||||||||||||||
Db       1984 AGLIPQAYGPRLHENAAADTEGYVLTGTSGSVASGRISYTFGFEGPAVSVDTACSSSLVA 2043

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2044 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 2103

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       2104 VGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 2163

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2164 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 2223

Qy        360 KMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 419
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2224 KMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 2283

Qy        420 NAHVILEEAPA 430
              ||||||||||||
Db       2284 NAHVILEEAPA 2294


RESULT 7
AAG65267
ID   AAG65267 standard; Protein; 5532 AA.
XX
AC   AAG65267;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 7.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 217-235; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   5532 AA;

  Query Match             86.9%;  Score 1968.5;  DB 22;  Length 5532;
  Best Local Similarity   87.2%;  Pred. No. 5.7e-166;
  Matches  376;  Conservative   23;  Mismatches   31;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              ||:||:|||||:|||||||:: |:|::| :| :| ||||||||: |:||:|  ||| ||
Db       1864 EPVAIIGMACRYPGGVTSAEELWELLASGRDTVGEFPTDRGWDLEALFDPEPGRPGTSYT 1923

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | |||||| ||| ||||||||||||||||||||:|| :| |||:| |: |: ||||
Db       1924 RCGSFLYDAGEFDAGFFGISPREALAMDPQQRLLLEASWEAMEQAGIDPTTVRGSQTGVF 1983

Qy        121 TGTNGQDYALRVH-NAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 179
               |   | |  |:| ||   |:|: ||||:||| ||||||||||||||||||||||||
Db       1984 AGLIPQAYGPRLHENAAADTEGYVLTGTSGSVASGRISYTFGFEGPAVSVDTACSSSLVA 2043

Qy        180 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2044 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 2103

Qy        240 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 299
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       2104 VGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 2163

Qy        300 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2164 GDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVI 2223

Qy        360 KMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 419
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2224 KMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGT 2283

Qy        420 NAHVILEEAPA 430
              ||||||||||||
Db       2284 NAHVILEEAPA 2294


RESULT 8
AAB23751
ID   AAB23751 standard; Protein; 4881 AA.
XX
AC   AAB23751;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:5.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
```

```
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 254-278; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   4881 AA;

  Query Match             77.0%;  Score 1744.5;  DB 21;  Length 4881;
  Best Local Similarity   76.7%;  Pred. No. 5.3e-146;
  Matches  329;  Conservative   40;  Mismatches   59;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||||||| ||:|||:||:| :||:| || |||||::  |||:|   || ||
Db       2636 EPIAIVGMACRFPGGVESAEDFWELIASGRDAVGEFPVDRGWDVEAFYDPEPGRAGTSYT 2695

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | ||||  |  ||| ||||||||||||||||||:|| :|| :| |||:| ||||: ||||
Db       2696 RCGGFLQGAAEFDAGFFGISPREALAMDPQQRLMLEVSWEALERAGIDPATLHGSTTGVF 2755

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              | : ||||  :    |  :|:|||| : ||:|||:||||||||||:||||||||||||
Db       2756 AGVSQQDYAELLRRGTQDHEGYALTGVSNSVVSGRLSYTFGFEGPAVTVDTACSSSLVAL 2815

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||:||||:||||||||||:|||||||||||||: || |||: : ||| || |||
Db       2816 HLACQALRSGECSLALAGGVTVMSTPGAFVEFSRQRGLSPDGRCKAYGSGADGVGWAEGV 2875

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||:| | ||||||||||||||||||||||||||||||||||||||||| ||||
Db       2876 GVLLVERLSEAERRGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALACAGLSVA 2935

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| || |||||||||||||||||||||| |:|| |:|||||||:|| ||||||||||
Db       2936 DVDVVEGHGTGTTLGDPIEAQALLATYGQGRSGERPVWLGSVKSNIGHAQAAAGVAGVIK 2995

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||| | ||| :||:||||||:|||:| |:|||| |||    || ||||||:|||||||
Db       2996 MVMALNHELLPTSLHIDEPSPHIDWSSGGVRLLTEPVPWQ-QNGRPRRAGVSAFGVSGTN 3054

Qy        421 AHVILEEAP 429
              ||||:|:||
Db       3055 AHVIIEQAP 3063


RESULT 9
AAG65266
ID   AAG65266 standard; Protein; 4881 AA.
XX
AC   AAG65266;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 6.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 201-217; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   4881 AA;

  Query Match             77.0%;  Score 1744.5;  DB 22;  Length 4881;
  Best Local Similarity   76.7%;  Pred. No. 5.3e-146;
  Matches  329;  Conservative   40;  Mismatches   59;  Indels    1;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||||||||||||| ||:|||:||:| :||:| || |||||::  |||:|   || ||
Db       2636 EPIAIVGMACRFPGGVESAEDFWELIASGRDAVGEFPVDRGWDVEAFYDPEPGRAGTSYT 2695

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | ||||  |  ||| ||||||||||||||||||:|| :|| :| |||:| ||||: ||||
Db       2696 RCGGFLQGAAEFDAGFFGISPREALAMDPQQRLMLEVSWEALERAGIDPATLHGSTTGVF 2755

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              | : ||||  :    |  :|:|||| : ||:|||:||||||||||:||||||||||||
Db       2756 AGVSQQDYAELLRRGTQDHEGYALTGVSNSVVSGRLSYTFGFEGPAVTVDTACSSSLVAL 2815

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||||||||:||||:||||||||||:|||||||||||||: || |||: : ||| || |||
Db       2816 HLACQALRSGECSLALAGGVTVMSTPGAFVEFSRQRGLSPDGRCKAYGSGADGVGWAEGV 2875

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|||||||:| | ||||||||||||||||||||||||||||||||||||||||| ||||
```

```
Db       2876 GVLLVERLSEAERRGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALACAGLSVA 2935

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| || |||||||||||||||||||||| |:|| |:||||||||:|| ||||||||||
Db       2936 DVDVVEGHGTGTTLGDPIEAQALLATYGQGRSGERPVWLGSVKSNIGHAQAAAGVAGVIK 2995

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||| | ||| :||:||||||:|||:| |:|||| |||    || |||||:|||||||
Db       2996 MVMALNHELLPTSLHIDEPSPHIDWSSGGVRLLTEPVPWQ-QNGRPRRAGVSAFGVSGTN 3054

Qy        421 AHVILEEAP 429
              ||||:|:||
Db       3055 AHVIIEQAP 3063


RESULT 10
AAY39299
ID   AAY39299 standard; Protein; 3170 AA.
XX
AC   AAY39299;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnC a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..423
FT                   /label= KS3
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 3"
FT   Domain          531..850
FT                   /label= AT3
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 3"
FT   Domain          1159..1337
FT                   /label= KR3
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 3"
FT   Domain          1425..1506
FT                   /label= ACP3
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 3"
FT   Domain          1529..1952
FT                   /label= KS4
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 4"
FT   Domain          2066..2396
FT                   /label= AT4
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 4"
FT   Domain          2700..2880
FT                   /label= KR4
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 4"
FT   Domain          2972..3053
FT                   /label= ACP4
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 4"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 103-113; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnC gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnC gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   3170 AA;

  Query Match             75.2%;  Score 1704;  DB 20;  Length 3170;
  Best Local Similarity   75.3%;  Pred. No. 1.2e-142;
  Matches  323;  Conservative   38;  Mismatches   66;  Indels    2;  Gaps    1;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| | ||:|||| :  | |::| |||| || |||| || ||||||| ||||||
Db         34 EPIAIVAMGCRYPGGVQDPEGLWKLVASGGDAIGEFPADRGWHLDELYDPDPDQPGTCYT 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| ||| || |||||||||||||||||| :|||:| ||::| :| |: ||||
Db         94 RHGGFLHDAGEFDAGFFDISPREALAMDPQQRLLLEISWETVESAGMDPRSLRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |   : |    | ||:  :|:  || |||| |||::| |||||||:|||||||||||
Db        154 AGLMYEGYDTGAHRAGEGVEGYLGTGNAGSVASGRVAYAFGFEGPAVTVDTACSSSLVAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:|| ||| :|||||||||:|  |||||||||| || ||:|:|||||||||||
Db        214 HLACQSLRQGECDLALAGGVTVMSTPERFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGA 273
```

```
Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |::|:|||||| |||||||||||||||||||||||||||| :|:|||:| | :|||||
Db        274 GLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGLAQERVIQQVLTSAGLSAS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| |||||||||:| |||||  : ||||||||||:|||||||||||||||
Db        334 DVDAVEAHGTGTRLGDPIEAQALIAAYGQDRDRDRPLWLGSVKSNIGHTQAAAGVAGVIK 393

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:||| |||||||||: |||||||||:|||| :?|||     ||||||||||||||
Db        394 MVMAMRHGELPRTLHVDEPNSHVDWSAGAVRLLTENIRWPG--TGTRRAGVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              |||||| |
Db        452 AHVILEHDP 460
```

RESULT 11
AAB70967
ID AAB70967 standard; Protein; 3170 AA.
XX
AC AAB70967;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa protein fragment encoded by ORF20, SEQ ID 46.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR N-PSDB; AAF88337.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 57; Page 226-238; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ Sequence 3170 AA;

Query Match 75.2%; Score 1704; DB 22; Length 3170;
Best Local Similarity 75.3%; Pred. No. 1.2e-142;
Matches 323; Conservative 38; Mismatches 66; Indels 2; Gaps 1;

```
Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              |||||| | ||:|||| : | |::| |||| || |||| || ||||||| ||||||
Db         34 EPIAIVAMGCRYPGGVQDPEGLWKLVASGGDAIGEFPADRGWHLDELYDPDPDQPGTCYT 93

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||:||| ||| || |||||||||||||||| :|||:| ||::| :| |: ||||
Db         94 RHGGFLHDAGEFDAGFFDISPREALAMDPQQRLLLEISWETVESAGMDPRSLRGSRTGVF 153

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
              |   : |    | ||:  :|:  || |||| |||::| |||||||:|||||||||||
Db        154 AGLMYEGYDTGAHRAGEGVEGYLGTGNAGSVASGRVAYAFGFEGPAVTVDTACSSSLVAL 213

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              |||||:|| ||| :||||||||||:|  |||||||||| || ||:|:|||||||||||
Db        214 HLACQSLRQGECDLALAGGVTVMSTPERFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGA 273

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |::|:|||||| |||||||||||||||||||||||||||| :|:|||:| | :|||||
Db        274 GLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGLAQERVIQQVLTSAGLSAS 333

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              |||||||||||| |||||||||:| |||||  : ||||||||||:|||||||||||||||
Db        334 DVDAVEAHGTGTRLGDPIEAQALIAAYGQDRDRDRPLWLGSVKSNIGHTQAAAGVAGVIK 393

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:||| |||||||||: |||||||||:|||| : |||     ||||||||||||||
Db        394 MVMAMRHGELPRTLHVDEPNSHVDWSAGAVRLLTENIRWPG--TGTRRAGVSSFGVSGTN 451

Qy        421 AHVILEEAP 429
              |||||| |
Db        452 AHVILEHDP 460
```

RESULT 12
AAB70968
ID AAB70968 standard; Protein; 4924 AA.
XX
AC AAB70968;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa protein fragment encoded by ORF21, SEQ ID 48.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR N-PSDB; AAF88338.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 57; Page 264-284; 354pp; German.

```
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   4924 AA;

  Query Match             73.8%;  Score 1672;  DB 22;  Length 4924;
  Best Local Similarity   75.2%;  Pred. No. 1.6e-139;
  Matches  322;  Conservative   34;  Mismatches   70;  Indels    2;  Gaps    2;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||||| |||||||:| :: | |::   ||: ||| ||||||  |||||||  || |
Db       1783 VAIVGMGCRFPGGVSSPEELWRLVAGGVDAVAGFPDDRGWDLAALYDPDPDRLGTSYVCE 1842

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTG 122
              |||| ||  |||: ||||||||||||||||||| ||||:| |||:| :|||: |||| |
Db       1843 GGFLRDAAEFDADMFGISPREALAMDPQQRLLLEVAWETLERAGIDPFSLHGSRTGVFAG 1902

Qy        123 TNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
                  ||  | :  | :  :|   || ||||:|||::|:||||||||:||||||||||||
Db       1903 LMYHDYGARFITRAPEGFEGHLGTGNAGSVLSGRVAYSFGFEGPAVTVDTACSSSLVALH 1962

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              || |||||||  ||||||||||:|  |||||||||| || ||:|:|||||||||| |
Db       1963 LAGQALRAGECEFALAGGVTVMSTPTTFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGAG 2022

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
              ::|:|||||| |||| ||||||||||||||||||||||||||||| ||| :||||  |
Db       2023 LVLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALTSAGLSVSD 2082

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              |||||||||| |||||||||:||||:||    |||||||||:|||||||||||||||
Db       2083 VDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAGVIKM 2142

Qy        362 VMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNA 421
              |||:| | |||||||||||  |||||| ||||||  |||   |||||||||||:|||||
Db       2143 VMAMRQGELPRTLHVDEPSAQVDWSAGTVQLLTENTPWP-DSGRLRRAGVSSFGISGTNA 2201

Qy        422 HVILEEAP 429
              |:|||: |
Db       2202 HLILEQPP 2209


RESULT 13
AAY39300
ID   AAY39300 standard; Protein; 4928 AA.
XX
AC   AAY39300;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnD a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..424
FT                   /label= KS5
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 5"
FT   Domain          539..866
FT                   /label= AT5
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 5"
FT   Domain          893..1078
FT                   /label= DH5
FT                   /note= "Dehydratase domain: part of extender
FT                         module 5"
FT   Domain          1384..1565
FT                   /label= KR5
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 5"
FT   Domain          1645..1726
FT                   /label= ACP5
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 5"
FT   Domain          1748..2172
FT                   /label= KS6
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 6"
FT   Domain          2283..2613
FT                   /label= AT6
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 6"
FT   Domain          2916..3095
FT                   /label= KR6
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 6"
FT   Domain          3188..3269
FT                   /label= ACP6
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 6"
FT   Domain          3291..3713
FT                   /label= KS7
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 7"
FT   Domain          3825..4153
FT                   /label= AT7
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 7"
FT   Domain          4344..4638
FT                   /label= KR7
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 7"
FT   Domain          4725..4806
FT                   /label= ACP7
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 7"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
```

```
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds -
XX
PS   Claim 1; Page 113-129; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnD gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnD gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   4928 AA;

  Query Match           73.8%;  Score 1672;  DB 20;  Length 4928;
  Best Local Similarity   75.2%;  Pred. No. 1.6e-139;
  Matches  322;  Conservative   34;  Mismatches   70;  Indels    2;  Gaps    2;

Qy          3 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 62
              :||||| |||||||:| :: | |::   ||: ||| ||||||  |||||||  || |
Db       1787 VAIVGMGCRFPGGVSSPEELWRLVAGGVDAVAGFPDDRGWDLAALYDPDPDRLGTSYVCE 1846

Qy         63 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTG 122
              |||| ||  |||: |||||||||||||||||||| ||||:| |||:| :|||: |||| |
Db       1847 GGFLRDAAEFDADMFGISPREALAMDPQQRLLLEVAWETLERAGIDPFSLHGSRTGVFAG 1906

Qy        123 TNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALH 181
                  ||  | :  | :  :|   || ||||:|||::|:|||||||:||||||||||||
Db       1907 LMYHDYGARFITRAPEGFEGHLGTGNAGSVLSGRVAYSFGFEGPAVTVDTACSSSLVALH 1966

Qy        182 LACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVG 241
              || |||||||  ||||||||||:|  |||||||||| || ||:|:|||||||||| |
Db       1967 LAGQALRAGECEFALAGGVTVMSTPTTFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGAG 2026

Qy        242 MLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGD 301
              ::|:|||||| |||| ||||||||||||||||||||||||||| ||| :||||  |
Db       2027 LVLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALTSAGLSVSD 2086

Qy        302 VDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKM 361
              |||||||||| ||||||||:||||:||    ||||||||||:|||||||||||||||
Db       2087 VDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAGVIKM 2146

Qy        362 VMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNA 421
              |||:| | |||||||||||  |||||| ||||||  |||   ||||||||||||:|||||
Db       2147 VMAMRQGELPRTLHVDEPSAQVDWSAGTVQLLTENTPWP-DSGRLRRAGVSSFGISGTNA 2205

Qy        422 HVILEEAP 429
                 |:|||: |
Db       2206 HLILEQPP 2213


RESULT 14
AAE10129
ID   AAE10129 standard; Protein; 11096 AA.
XX
AC   AAE10129;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysC.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          35..455
FT                   /label= KS3_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          546..858
FT                   /label= AT3_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          872..1073
FT                   /label= DH3_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1381..1628
FT                   /label= KR3_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1662..1735
FT                   /label= ACP3_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1757..2180
FT                   /label= KS4_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2291..2603
FT                   /label= AT4_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2617..2818
FT                   /label= DH4_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3124..3371
FT                   /label= KR4_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3407..3480
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3501..3924
FT                   /label= KS5_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4032..4346
FT                   /label= AT5_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4360..4561
FT                   /label= DH5_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4953..5239
FT                   /label= ER5_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          5248..5495
FT                   /label= KR5_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5528..5601
FT                   /label= ACP5_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          5623..6046
FT                   /label= KS6_domain
FT                   /note= "Ketosynthase (KS) domain"
```

```
FT   Domain          6166..6478
FT                   /label= AT6_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          6492..6704
FT                   /label= DH6_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          7038..7281
FT                   /label= KR6_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          7315..7388
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          7408..7831
FT                   /label= KS7_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          7939..8253
FT                   /label= AT7_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          8267..8470
FT                   /label= DH7_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          8812..9086
FT                   /label= KR7_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          9120..9193
FT                   /label= ACP7_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          9214..9637
FT                   /label= KS8_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          9758..10072
FT                   /label= AT8_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          10086..10289
FT                   /label= DH8_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          10657..10904
FT                   /label= KR8_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          10939..11012
FT                   /label= ACP8_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17184.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 170-176; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin
CC   gene, NysC.
XX
SQ   Sequence   11096 AA;

  Query Match             73.1%;  Score 1656;  DB 22;  Length 11096;
  Best Local Similarity   73.6%;  Pred. No. 1.4e-137;
  Matches  320;  Conservative   40;  Mismatches   67;  Indels    8;  Gaps    3;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              :|| :|||||||||||| :| | |:    ||| |||:|||||| ||||||:| || ||
Db       3501 DPIVVVGMACRFPGGVTSPEDLWRLVDDGTDAITTFPTNRGWDLDNLYDPDPEHFGTSYT 3560

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              |:||||::|| ||  |||:|||||| | |||||||::|| || |||:| || |: ||||
Db       3561 RSGGFLHEAGEFDPAFFGMSPREALATDSQQRLLLESSWEAIERAGIDPLTLRGSATGVF 3620

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
               |    ||   :   |:  :||   |:|||| |||:||  |||||||:|||||||||||
Db       3621 AGVMYSDYGSIL--GGKEFEGFQGQGSAGSVASGRVSYALGFEGPAVTVDTACSSSLVAL 3678

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              | | |||||||||:||||||||||:|  |||||||||| ||  |||: |||| || |||
Db       3679 HWAAQALRAGECSLALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKAFAEAADGVGWSEGV 3738

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
              |:|::|| ||| |||| :|||:|||||||||||||||||||||||||||||||||: |||
Db       3739 GILVLERQSDAVRNGHEILAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSTA 3798

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||||||||||||||||||||||||||||:||  | ||||||:|||:|||||||||||||
Db       3799 DVDAVEAHGTGTTLGDPIEAQALLATYGRDRDPENPLWLGSLKSNIGHTQAAAGVAGVIK 3858

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||||:|||:||:||||| || ||||| |||:||||   ||   ||:|||||||||:||||
Db       3859 MVMAMRHGVLPQTLHVDAPSSHVDWSVGAVELLTEQTAWP-ETGRVRRAGVSSFGISGTN 3917

Qy        421 AHVILE-----EAPA 430
              ||||:|     |:||
Db       3918 AHVIVEQPALVESPA 3932


RESULT 15
AAY92709
ID   AAY92709 standard; Protein; 3519 AA.
XX
AC   AAY92709;
XX
DT   29-AUG-2000  (first entry)
XX
DE   S. antibioticus 8,8a-deoxyoleandolide synthase III encoded by oleAIII.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   Region          33..1564
FT                   /label= Extender_module_5
FT   Domain          33..466
FT                   /label= KS5_domain
FT                   /note= "ketosynthase domain"
FT   Domain          569..915
FT                   /label= AT5_domain
```

```
FT                   /note= "acyl transferase domain"
FT   Domain          1199..1383
FT                   /label= KR5_domain
FT                   /note= "ketoreductase domain"
FT   Domain          1478..1564
FT                   /label= ACP5_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          1686..3218
FT                   /label= Extender_module_6
FT   Domain          1686..2115
FT                   /label= KS6_domain
FT                   /note= "ketosynthase domain"
FT   Domain          2220..2566
FT                   /label= AT6_domain
FT                   /note= "acyl transferase domain"
FT   Domain          2855..3039
FT                   /label= KR6_domain
FT                   /note= "ketoreductase domain"
FT   Domain          3132..3218
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein domain"
FT   Domain          3290..3519
FT                   /label= TE_domain
FT                   /note= "thiol esterase domain"
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;    99WO-US24478.
XX
PR   29-OCT-1998;    98US-0106100.
PR   16-FEB-1999;    99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   N-PSDB; AAA09469.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 29; 86pp; English.
XX
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence   3519 AA;

  Query Match             73.0%;  Score 1654;  DB 21;  Length 3519;
  Best Local Similarity   73.3%;  Pred. No. 4.2e-138;
  Matches  315;  Conservative   40;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          1 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 60
              | |||  |:||| ||: | :| |  ::  :||: | | ||||||| || |||::||| |
Db       1687 EAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGTTYV 1746

Qy         61 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 120
              | | | |||  ||| ||||||||||||||||||||||:||  | | |:|:|: || ||:|
Db       1747 REGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTATGIF 1806

Qy        121 TGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVAL 180
                |   | |       | :|  |: ||||| :|:|||||||| |||||:||||||||||
Db       1807 IGAGHQGYGPDPKRAPESVAGYLLTGTASAVLSGRISYTFGLEGPAVTVDTACSSSLVAL 1866

Qy        181 HLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGV 240
              ||| |||| ||||:|:|||| |||:| |||||||:|:| || ||||:|||| ||||||
Db       1867 HLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAADGMGWGEGV 1926

Qy        241 GMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAG 300
               :||:|||||| | ||||||||||||||||||| |||||||||| |||:|||:
Db       1927 SLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPA 1986

Qy        301 DVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIK 360
              ||| |||||||| |||||||||||||||| |||  |:||||||||:|||||||||||:|
Db       1987 DVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAAAGVAGVMK 2046

Qy        361 MVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTN 420
              ||:||  |::|:||||||||||||||||:|||| ||      ||||||:|:|||||||
Db       2047 MVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGISAFGVSGTN 2106

Qy        421 AHVILEEAPA 430
              ||||:|||||
Db       2107 AHVIVEEAPA 2116


  Search completed: June 17, 2003, 13:02:32
  Job time : 22.0646 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 3.78986 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_1979_2060
Perfect score: 407
Sequence: 1 LAHQTSADQRAALLELVRDH.....RLPVSLVFDHPTPAKLAVHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	407	100.0	3626	2	Q9F7T9	Q9f7t9 streptomyce
2	407	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
3	243	59.7	5532	2	Q9S0R4	Q9s0r4 streptomyce
4	240.5	59.1	3524	2	Q93H86	Q93h86 streptomyce
5	240	59.0	6797	2	Q9X993	Q9x993 streptomyce
6	238	58.5	3576	16	Q9EX53	Q9ex53 streptomyce
7	233	57.2	3939	2	Q93HJ3	Q93hj3 streptomyce
8	233	57.2	6239	2	Q9S0R7	Q9s0r7 streptomyce
9	229.5	56.4	3352	2	Q93H83	Q93h83 streptomyce
10	229	56.3	1839	2	O30765	O30765 streptomyce
11	229	56.3	3739	2	Q9ZGI4	Q9zgi4 streptomyce
12	229	56.3	3970	2	Q93HI8	Q93hi8 streptomyce
13	229	56.3	6048	2	Q93H87	Q93h87 streptomyce
14	229	56.3	9477	2	Q9L4X3	Q9l4x3 streptomyce
15	226	55.5	6146	2	Q93HJ5	Q93hj5 streptomyce
16	225.5	55.4	3654	2	O30766	O30766 streptomyce
17	225	55.3	1562	2	Q9ZGI3	Q9zgi3 streptomyce
18	224	55.0	1762	2	O52546	O52546 amycolatops
19	224	55.0	1763	2	O52790	O52790 amycolatops
20	224	55.0	3816	2	Q9KIV3	Q9kiv3 streptomyce
21	223	54.8	2100	2	O30481	O30481 streptomyce
22	223	54.8	4150	2	Q9KIV4	Q9kiv4 streptomyce
23	223	54.8	9510	2	Q93NX9	Q93nx9 streptomyce
24	222	54.5	6396	2	Q9KID7	Q9kid7 streptomyce
25	221	54.3	5060	2	O52545	O52545 amycolatops
26	221	54.3	5069	2	O52789	O52789 amycolatops
27	221	54.3	6420	2	P95814	P95814 streptomyce
28	221	54.3	11096	2	Q9L4W3	Q9l4w3 streptomyce
29	220	54.1	4840	2	Q93HJ4	Q93hj4 streptomyce
30	219	53.8	3729	2	O33956	O33956 streptomyce
31	219	53.8	4685	2	Q93HJ2	Q93hj2 streptomyce
32	218	53.6	3613	2	Q93HJ1	Q93hj1 streptomyce
33	217	53.3	4881	2	Q9S0R3	Q9s0r3 streptomyce
34	216.5	53.2	1762	2	O30480	O30480 streptomyce
35	216	53.1	3413	2	O54593	O54593 amycolatops
36	216	53.1	3591	2	Q9KIE1	Q9kie1 streptomyce
37	216	53.1	4151	16	O53490	O53490 mycobacteri
38	215.5	52.9	7525	2	Q9KIE0	Q9kie0 streptomyce
39	215	52.8	5435	2	Q9L4X2	Q9l4x2 streptomyce
40	215	52.8	8563	2	Q54297	Q54297 streptomyce
41	214.5	52.7	10223	2	Q54296	Q54296 streptomyce
42	212	52.1	1728	2	O54591	O54591 amycolatops
43	212	52.1	2152	2	Q9ALM5	Q9alm5 saccharopol
44	212	52.1	6260	2	Q54299	Q54299 streptomyce
45	211	51.8	10917	2	Q93NW6	Q93nw6 streptomyce

ALIGNMENTS

RESULT 1
Q9F7T9
ID Q9F7T9 PRELIMINARY; PRT; 3626 AA.
AC Q9F7T9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Avermectin polyketide synthase (Fragment).
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31271;
RA Hong Y.-S., Lee J.J.;
RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT and polyketide synthase gene from Streptomyces avermitilis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275943; AAG09812.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT NON_TER 3626 3626
SQ SEQUENCE 3626 AA; 380557 MW; 6272F5F088C1A8D0 CRC64;

Query Match 100.0%; Score 407; DB 2; Length 3626;
Best Local Similarity 100.0%; Pred. No. 1.6e-34;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db       1946 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 2005

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              |||||||||||||||||||||
Db       2006 GLRLPVSLVFDHPTPAKLAVHL 2027


RESULT 2
Q9S0R8
ID   Q9S0R8      PRELIMINARY;      PRT;  3972 AA.
AC   Q9S0R8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 1.
GN   AVEA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84474.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3972 AA;  416852 MW;  2A293695B032B1C3 CRC64;

  Query Match             100.0%;  Score 407;  DB 2;  Length 3972;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-34;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 2038

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||||||||||||||||||||
Db       2039 GLRLPVSLVFDHPTPAKLAVHL 2060


RESULT 3
Q9S0R4
ID   Q9S0R4      PRELIMINARY;      PRT;  5532 AA.
AC   Q9S0R4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 3.
GN   AVEA3.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84478.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5532 AA;  575193 MW;  EC7B2069D44A4057 CRC64;

  Query Match             59.7%;  Score 243;  DB 2;  Length 5532;
  Best Local Similarity   59.8%;  Pred. No. 1.1e-16;
  Matches   49;  Conservative    5;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || ||   |   || ||| |:| || |   |  | ||::|| |||||||||| || | : |
Db       1756 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 1815

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :| |||| |  |  ||
Db       1816 GLRLPTTLAFDHPNPTTLTHHL 1837


RESULT 4
Q93H86
ID   Q93H86      PRELIMINARY;      PRT;  3524 AA.
AC   Q93H86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69304.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3524 AA;  366619 MW;  64A65759C461EBC4 CRC64;

  Query Match             59.1%;  Score 240.5;  DB 2;  Length 3524;
  Best Local Similarity   65.4%;  Pred. No. 1.2e-16;
```

```
  Matches   51;  Conservative    9;  Mismatches   17;  Indels    1;  Gaps    1;

Qy          6 SADQR-AALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 64
              :||:|  |||||||  ||||| |:  :|: |  :|: |||||||||| || |  ||||||
Db       1627 AADERDEALLELVRGEVAAVLGHSSSRAVQPTYAFQDLGFDSLTAVELRNRLATATGLRL 1686

Qy         65 PVSLVFDHPTPAKLAVHL 82
              | :|||||||| || |:
Db       1687 PATLVFDHPTPAALARHI 1704


RESULT 5
Q9X993
ID   Q9X993      PRELIMINARY;      PRT;  6797 AA.
AC   Q9X993;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase (PIMS1 protein).
GN   PIMS1.
OS   Streptomyces natalensis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=68242;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99214571; PubMed=10187796;
RA   Aparicio J.F., Colina A.J., Ceballos E., Martin J.F.;
RT   "The biosynthetic gene cluster for the 26-membered ring polyene
RT   macrolide pimaricin. A new polyketide synthase organization encoded by
RT   two subclusters separated by functionalization genes.";
RL   J. Biol. Chem. 274:10133-10139(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Aparicio J.F., Fouces R., Mendes M.V., Olivera N.;
RT   "A complex multienzyme system encoded by five polyketide synthase
RT   genes is involved in the biosynthesis of the 26-membered polyene
RT   macrolide pimaricin in Streptomyces natalensis.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ132222; CAB41041.1; -.
DR   EMBL; AJ278573; CAC20931.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00626; RCC1_2; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6797 AA;  710149 MW;  1D2DF85074D050DA CRC64;

  Query Match             59.0%;  Score 240;  DB 2;  Length 6797;
  Best Local Similarity   59.8%;  Pred. No. 2.8e-16;
  Matches   49;  Conservative   14;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  ::|::| |||::||  :| || || |: ||||::|: || |||||:| |||| |||
Db       4893 LAGLSTAERREALLDVVRAQIATVLGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKAT 4952

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| : |||:||   || ||
Db       4953 GLRLPATTVFDYPTVDALAAHL 4974


RESULT 6
Q9EX53
ID   Q9EX53      PRELIMINARY;      PRT;  3576 AA.
AC   Q9EX53;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative type I polyketide synthase.
GN   SCO6274 OR SC2C4.04C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL512902; CAC22145.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR003662; sub_transporter.
DR   InterPro; IPR001998; Xylose_isom.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   TIGRFAMs; TIGR00128; fabD; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR   PROSITE; PS00173; XYLOSE_ISOMERASE_2; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3576 AA;  377350 MW;  8A83C6F90AC9928B CRC64;

  Query Match             58.5%;  Score 238;  DB 16;  Length 3576;
  Best Local Similarity   61.0%;  Pred. No. 2.2e-16;
  Matches   50;  Conservative   12;  Mismatches   20;  Indels    0;  Gaps    0;
```

```
Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  ::|::  || :|||  ||||| |:|| ||   ::|: |||||||||| || |  |:
Db       1688 LAPLSAAERERALTDLVRVQVAAVLGHSDPGAIESGRAFQELGFDSLTAVELRNQLSTAS 1747

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :||||||:|| || ||
Db       1748 GLRLPTTLVFDHPSPAALAAHL 1769


RESULT 7
Q93HJ3
ID   Q93HJ3      PRELIMINARY;      PRT;  3939 AA.
AC   Q93HJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA3.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69194.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3939 AA;  410482 MW;  534AE3B517256503 CRC64;

 Query Match             57.2%;  Score 233;  DB 2;  Length 3939;
  Best Local Similarity   61.3%;  Pred. No. 8.4e-16;
  Matches   46;  Conservative    9;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          5 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 64
              |  ::   ||:||| ||| || |:|| |:     || ||||||||||||| |  | | ||
Db       2009 TGTERMRILLQLVRSHVATVLGHSDPDAVGSTAPFRELGFDSLTAVEFRNRLTGAVGFRL 2068

Qy         65 PVSLVFDHPTPAKLA 79
              ||::|||||||::||
Db       2069 PVTVVFDHPTPSELA 2083


RESULT 8
Q9S0R7
ID   Q9S0R7      PRELIMINARY;      PRT;  6239 AA.
AC   Q9S0R7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 2.
GN   AVEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84475.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6239 AA;  666283 MW;  A706280940B502FA CRC64;

  Query Match             57.2%;  Score 233;  DB 2;  Length 6239;
  Best Local Similarity   61.0%;  Pred. No. 1.4e-15;
  Matches   50;  Conservative    8;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || |:: ::| ||| ||| |||||| |:    |   ::|| |||||||||| || |  ||
Db       6032 LARQSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAAT 6091

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||  :| || |||| || ||
Db       6092 GLRLRATLAFDFPTPAALAEHL 6113


RESULT 9
Q93H83
ID   Q93H83      PRELIMINARY;      PRT;  3352 AA.
AC   Q93H83;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA5.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69307.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002362; Antenna_beta.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
```

```
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00969; ANTENNA_COMP_BETA; UNKNOWN_1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3352 AA;  353395 MW;  66A5865BFD2027B7 CRC64;

  Query Match             56.4%;  Score 229.5;  DB 2;  Length 3352;
  Best Local Similarity   58.5%;  Pred. No. 1.7e-15;
  Matches   48;  Conservative   10;  Mismatches   21;  Indels    3;  Gaps    1;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              |:||   |:  ||::||| | |||| |     |:  |: || :||||||||| || : |||
Db       1426 LSHQ---DRARALVDLVRTHAAAVLGHRGAGAVEADRPFRDIGFDSLTAVELRNRINKAT 1482

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              || || :|:||||| | || ||
Db       1483 GLSLPATLIFDHPTAADLAAHL 1504


RESULT 10
O30765
ID   O30765      PRELIMINARY;      PRT;  1839 AA.
AC   O30765;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase module 3.
GN   NIDA2.
OS   Streptomyces caelestis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=36816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL-2821;
RX   MEDLINE=98053867; PubMed=9393718;
RA   Kakavas S.J., Katz L., Stassi D.;
RT   "Identification and characterization of the niddamycin polyketide
RT   synthase genes from Streptomyces caelestis.";
RL   J. Bacteriol. 179:7515-7522(1997).
DR   EMBL; AF016585; AAC46025.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   1839 AA;  194158 MW;  46683C00526090A2 CRC64;

  Query Match             56.3%;  Score 229;  DB 2;  Length 1839;
  Best Local Similarity   60.8%;  Pred. No. 9.3e-16;
  Matches   48;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || :| |::   ||||||  ||||| | || |:  ::||:  ||||||||: || |   |
Db       1662 LARRTVAERLPLLLELVRTEVAAVLGHGDPGAVGAERSFKDAGFDSLTAVDLRNRLNART 1721

Qy         61 GLRLPVSLVFDHPTPAKLA 79
              ||||| :||||||||  ||
Db       1722 GLRLPATLVFDHPTPLALA 1740


RESULT 11
Q9ZGI4
ID   Q9ZGI4      PRELIMINARY;      PRT;  3739 AA.
AC   Q9ZGI4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAII.
GN   PIKAII.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR   EMBL; AF079138; AAC69330.1; -.
DR   HSSP; Q02054; 2AF8.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3739 AA;  387170 MW;  3D3910824DA5B080 CRC64;

  Query Match             56.3%;  Score 229;  DB 2;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 2.1e-15;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 12
Q93HI8
ID   Q93HI8      PRELIMINARY;      PRT;  3970 AA.
AC   Q93HI8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
```

```
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69199.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3970 AA;  412657 MW;  A93E583FAFA68C47 CRC64;

  Query Match             56.3%;  Score 229;  DB 2;  Length 3970;
  Best Local Similarity   65.8%;  Pred. No. 2.3e-15;
  Matches   50;  Conservative    6;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          7 ADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPV 66
              |:|   ||:||| : | || ||   ||| :|||  |||||||||||||| |  |||||||
Db       3811 AEQLRRLLDLVRSNAATVLGHASAGAIAAEQSFSELGFDSLTAVEFRNRLGAATGLRLPA 3870

Qy         67 SLVFDHPTPAKLAVHL 82
              :|||:||||| ||  |
Db       3871 TLVFEHPTPAALAGEL 3886


RESULT 13
Q93H87
ID   Q93H87      PRELIMINARY;      PRT;  6048 AA.
AC   Q93H87;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69303.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
KW   Phosphopantetheine.
SQ   SEQUENCE   6048 AA;  629420 MW;  F8673C2ED2694705 CRC64;

  Query Match             56.3%;  Score 229;  DB 2;  Length 6048;
  Best Local Similarity   57.3%;  Pred. No. 3.7e-15;
  Matches   47;  Conservative   11;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              |     |::   :|:|||  ||| | ||  : : |: :|: |||||||||| || | |||
Db       5878 LEEMPEAEREKTVLDLVRAEVAAGLGHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKAT 5937

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :||||||||| || |:
Db       5938 GLRLPATLVFDHPTPALLARHV 5959


RESULT 14
Q9L4X3
ID   Q9L4X3      PRELIMINARY;      PRT;  9477 AA.
AC   Q9L4X3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysI.
GN   NYSI.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71766.1; -.
DR   HSSP; P81989; 1QE5.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   9477 AA;  988132 MW;  0D2BCA5D7B265483 CRC64;

  Query Match             56.3%;  Score 229;  DB 2;  Length 9477;
  Best Local Similarity   59.8%;  Pred. No. 6.3e-15;
  Matches   49;  Conservative    5;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || | :||  | || || :  | || |||   : ||  || |||||||||: || |   |
Db       7801 LAEQPAADHDAILLALVTEKAATVLGHADHDLLEPDLPFRDLGFDSLTAVDLRNQLTAET 7860

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              || || :|||||| || || ||
Db       7861 GLTLPATLVFDHPNPAALAAHL 7882
```

```
RESULT 15
Q93HJ5
ID   Q93HJ5      PRELIMINARY;      PRT;  6146 AA.
AC   Q93HJ5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69192.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR002155; Thiolase.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF00108; thiolase; 1.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
DR   PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   6146 AA;  637460 MW;  1D56F1108D458D00 CRC64;

  Query Match             55.5%;  Score 226;  DB 2;  Length 6146;
  Best Local Similarity   56.1%;  Pred. No. 8e-15;
  Matches   46;  Conservative   10;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  : |::   :|:|||:| | || |     | |||:|: |||||||||| || |  ||
Db       5980 LAALSPAERHRLVLDLVREHTAGVLGHGSADDIDPDQAFKTLGFDSLTAVELRNRLRTAT 6039

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
                | :| :||||||||| |  ||
Db       6040 SLTVPATLVFDHPTPAALTAHL 6061


Search completed: June 17, 2003, 13:07:55
Job time : 4.78986 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 0.894236 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_1979_2060
Perfect score: 407
Sequence: 1 LAHQTSADQRAALLELVRDH.....RLPVSLVFDHPTPAKLAVHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	205	50.4	3172	1	ERY3_SACER	Q03133 saccharopol
2	205	50.4	3567	1	ERY2_SACER	Q03132 saccharopol
3	202	49.6	3491	1	ERY1_SACER	Q03131 saccharopol
4	186	45.7	3519	1	OL56_STRAT	Q07017 streptomyce
5	104.5	25.7	4427	1	PKSL_BACSU	Q05470 bacillus su
6	98	24.1	1538	1	PPSB_MYCTU	Q10978 mycobacteri
7	94.5	23.2	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	87	21.4	1278	1	DHBF_BACSU	P45745 bacillus su
9	82.5	20.3	1774	1	MSAS_PENPA	P22367 penicillium
10	77	18.9	2110	1	MCAS_MYCBO	Q02251 mycobacteri
11	75.5	18.6	1392	1	LYS2_YEAST	P07702 saccharomyc
12	71.5	17.6	2505	1	FAS_RAT	P12785 rattus norv
13	71	17.4	92	1	ACP_STRCO	Q9rk61 streptomyce
14	70.5	17.3	5217	1	HTS1_COCCA	Q01886 cochliobolu
15	70	17.2	3587	1	SRF1_BACSU	P27206 bacillus su
16	69	17.0	6486	1	TYCC_BACBR	O30409 b tyrocidin
17	68.5	16.8	838	1	FAS_MOUSE	P19096 mus musculu
18	68.5	16.8	3649	1	ACVS_NOCLA	P27743 nocardia la
19	68	16.7	5255	1	BACA_BACLI	O68006 b bacitraci
20	67.5	16.6	3770	1	ACVS_EMENI	P27742 emericella
21	66.5	16.3	1274	1	SRF3_BACSU	Q08787 bacillus su
22	66	16.2	2035	1	HMP2_YEREN	P48633 yersinia en
23	65	16.0	88	1	ACPX_STRCM	P41174 streptomyce
24	65	16.0	2560	1	PPS2_BACSU	P39846 bacillus su
25	65	16.0	3587	1	TYCB_BACBR	O30408 b tyrocidin
26	64.5	15.8	122	1	ACPM_ARATH	P53665 arabidopsis
27	64.5	15.8	357	1	DP42_AGRT5	Q8ujk7 agrobacteri
28	63.5	15.6	568	1	DNLI_METAC	Q8tmt1 methanosarc
29	63.5	15.6	612	1	CRY2_ARATH	Q96524 arabidopsis
30	63	15.5	385	1	YL90_MYCTU	Q10383 mycobacteri
31	62.5	15.4	332	1	GALR_HAEIN	P31766 haemophilus
32	62	15.2	295	1	ISPE_CAUCR	Q9a817 caulobacter
33	62	15.2	420	1	GATZ_ECOLI	P37191 escherichia
34	61.5	15.1	4447	1	PKSK_BACSU	P40803 bacillus su
35	61	15.0	242	1	HEM3_CHLMU	Q9pk95 chlamydia m
36	61	15.0	262	1	TRPC_AZOBR	P26938 azospirillu
37	61	15.0	1167	1	CARB_RHILO	Q98187 rhizobium l
38	60	14.7	77	1	ACP_SYNY3	P20804 synechocyst
39	60	14.7	601	1	HUTU_RALSO	Q8xw28 ralstonia s
40	59.5	14.6	4273	1	PKSM_BACSU	P40872 bacillus su
41	59	14.5	88	1	ACPM_BOVIN	P52505 bos taurus
42	59	14.5	156	1	ACPM_HUMAN	O14561 homo sapien
43	59	14.5	211	1	LEXA_XYLFA	Q9ph24 xylella fas
44	59	14.5	859	1	SUL1_YEAST	P38359 saccharomyc
45	59	14.5	3712	1	ACVS_CEPAC	P25464 cephalospor

ALIGNMENTS

RESULT 1
ERY3_SACER
ID ERY3_SACER STANDARD; PRT; 3172 AA.
AC Q03133; Q54097; Q99270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=91043075; PubMed=2234082;
RA Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-
RT producing polyketide synthase of Saccharopolyspora erythraea.";
RL Nature 348:176-178(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION

```
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match          50.4%;  Score 205;  DB 1;  Length 3172;
  Best Local Similarity   58.6%;  Pred. No. 3e-14;
  Matches   41;  Conservative    8;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         13 LLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDH 72
              |||    ||||:| |: | |:  || |  |||||||||  || | :|||| || :|||:|
Db       2821 LLEFTHSHVAAILGHSSPDAVGQDQPFTELGFDSLTAVGLRNQLQQATGLALPATLVFEH 2880

Qy         73 PTPAKLAVHL 82
              ||  :|| |:
Db       2881 PTVRRLADHI 2890


RESULT 2
ERY2_SACER
ID   ERY2_SACER     STANDARD;      PRT;  3567 AA.
AC   Q03132; Q54096;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE   deoxyerythronolide B synthase II) (DEBS 2).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
```

```
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match           50.4%;  Score 205;  DB 1;  Length 3567;
  Best Local Similarity  53.7%;  Pred. No. 3.4e-14;
  Matches   44;  Conservative   11;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || :: :|| | | |||| | |||  :     :   ::|: |||||| ||| || |  ||
Db       3403 LAGRSESDQVAGLAELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTAT 3462

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              |:||| :||||||||  :| ||
Db       3463 GVRLPSTLVFDHPTPLAVAEHL 3484


RESULT 3
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
```

```
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             49.6%;  Score 202;  DB 1;  Length 3491;
  Best Local Similarity   52.4%;  Pred. No. 7e-14;
  Matches   43;  Conservative   10;  Mismatches   29;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    : ::  || |||| | |||| ||  : :  ||:|  || |||:|:| || |  ||
Db       1876 LASLPAPEREEALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAAT 1935

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              |:||| : |||||    || ||
Db       1936 GVRLPTTTVFDHPDVRTLAAHL 1957


RESULT 4
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1     ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             45.7%;  Score 186;  DB 1;  Length 3519;
  Best Local Similarity   50.6%;  Pred. No. 4e-12;
  Matches   40;  Conservative   12;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              |:  : | : | |::|||   |||| |   : :  :::|: ||||||||: || |  ||
Db       3133 LSMLSPAGREAELVKLVRTEAAAVLGHGSAQDVPAERAFKELGFDSLTAVQLRNRLAAAT 3192

Qy         61 GLRLPVSLVFDHPTPAKLA 79
              | ||| | |||||  | ||
Db       3193 GTRLPASAVFDHPHAAALA 3211


RESULT 5
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
```

```
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;

  Query Match             25.7%;  Score 104.5;  DB 1;  Length 4427;
  Best Local Similarity   34.2%;  Pred. No. 0.0045;
  Matches   27;  Conservative   16;  Mismatches   35;  Indels    1;  Gaps    1;

Qy          4 QTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLR 63
              :|      || : :   :| || : |   | ||::|  :| ||:| :|:   : | |
Db       2482 ETETITAEALCDELTAGLAEVL-YMDQNEIDPDEAFIDIGMDSITGLEWIKAINKQYGTS 2540

Qy         64 LPVSLVFDHPTPAKLAVHL 82
              | |: |:|:||    ||:|
Db       2541 LNVTKVYDYPTTRDFAVYL 2559


RESULT 6
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             24.1%;  Score 98;  DB 1;  Length 1538;
  Best Local Similarity   34.7%;  Pred. No. 0.0075;
  Matches   25;  Conservative   13;  Mismatches   30;  Indels    4;  Gaps    1;

Qy         15 ELVRDHV----AAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVF 70
              ::: |||    | |:     : : |   |  || ||| :|  :  | :: |  || |:||
Db       1423 DMLFDHVGALAATVMGMPPTEPLDPSAGFFQLGMDSLMSVTLQRALSESLGEFLPASVVF 1482

Qy         71 DHPTPAKLAVHL 82
              |:||   |  :|
Db       1483 DYPTVYSLTDYL 1494


RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             23.2%;  Score 94.5;  DB 1;  Length 1876;
  Best Local Similarity   32.5%;  Pred. No. 0.023;
  Matches   25;  Conservative   17;  Mismatches   34;  Indels    1;  Gaps    1;

Qy          6 SADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLP 65
              : : |: | : :|  :|| ||    | :  |: |  || :|| |:  |    :  |: |
Db       1757 ATEVRSELEQGLRRIIAAELR-VPEKELDTDRPFAELGLNSLMAMAIRREAEQFVGIELS 1815

Qy         66 VSLVFDHPTPAKLAVHL 82
              :::|:|||   || :|
Db       1816 ATMLFNHPTVKSLASYL 1832


RESULT 8
DHBF_BACSU
ID   DHBF_BACSU     STANDARD;      PRT;  1278 AA.
AC   P45745;
DT   01-NOV-1995 (Rel. 32, Created)
```

```
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable serine activating enzyme.
GN   DHBF.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE OF 1-95 FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=96146538; PubMed=8550523;
RA   Rowland B.M., Taber H.W.;
RT   "Duplicate isochorismate synthase genes of Bacillus subtilis:
RT   regulation and involvement in the biosyntheses of menaquinone and
RT   2,3-dihydroxybenzoate.";
RL   J. Bacteriol. 178:854-861(1996).
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: 2,3-dihydroxybenzoate biosynthesis.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99120; CAB15186.1; -.
DR   EMBL; U26444; AAC44634.1; -.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG11243; dhbF.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00668; Condensation; 2.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Ligase; Phosphopantetheine; Complete proteome.
FT   DOMAIN      966   1033       ACYL CARRIER (ACP).
FT   BINDING     996    996       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1278 AA;  142068 MW;  B44FBFF3FCB085B4 CRC64;

  Query Match             21.4%;  Score 87;  DB 1;  Length 1278;
  Best Local Similarity   35.8%;  Pred. No. 0.099;
  Matches   29;  Conservative    9;  Mismatches   37;  Indels    6;  Gaps    2;

Qy          5 TSADQRAALL---ELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATG 61
              ||   ||      |:: |  | ||  |    :  | ||  ||  || |    : : :  |
Db        955 TSVSDRAPRTPQEEILCDLFAEVLGLA---RVGIDDSFFELGGHSLLAARLMSRIREVMG 1011

Qy         62 LRLPVSLVFDHPTPAKLAVHL 82
                | :: :|| || | || ||
Db       1012 AELGIAKLFDEPTVAGLAAHL 1032


RESULT 9
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
```

```
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             20.3%;  Score 82.5;  DB 1;  Length 1774;
  Best Local Similarity   36.0%;  Pred. No. 0.44;
  Matches   27;  Conservative    9;  Mismatches   30;  Indels    9;  Gaps    2;

Qy         10 RAALLELVRDHVAAVLR----HADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLP 65
              :| | | :|  || ||:      | ||    |      || ||:  |  |  |     : :|
Db       1697 KAYLDEKIRGCVAKVLQMTAEDVDSKAALAD-----LGVDSVMTVTLRRQLQLTLKIAVP 1751

Qy         66 VSLVFDHPTPAKLAV 80
                :| : ||| : |||
Db       1752 PTLTWSHPTVSHLAV 1766


RESULT 10
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             18.9%;  Score 77;  DB 1;  Length 2110;
  Best Local Similarity   31.4%;  Pred. No. 2.1;
  Matches   22;  Conservative   13;  Mismatches   31;  Indels    4;  Gaps    1;

Qy         13 LLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDH 72
              |  |: :  : :||    : |  |:||    | |||  :| |  :   ||:||   ::  :
Db       2030 LRRLLVEQASVILR----RTIDADRSFIEYGLDSLGMLEMRTHVETETGIRLTPKVIATN 2085

Qy         73 PTPAKLAVHL 82
                |   || :|
Db       2086 KTARALAQYL 2095


RESULT 11
LYS2_YEAST
ID   LYS2_YEAST     STANDARD;      PRT;  1392 AA.
AC   P07702;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aminoadipate-semialdehyde dehydrogenase large subunit (EC 1.2.1.31)
DE   (Alpha-aminoadipate reductase) (Alpha-AR).
GN   LYS2 OR YBR115C OR YBR0910.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91192607; PubMed=2013406;
RA   Morris M.E., Jinks-Robertson S.;
RT   "Nucleotide sequence of the LYS2 gene of Saccharomyces cerevisiae:
RT   homology to Bacillus brevis tyrocidine synthetase 1.";
RL   Gene 98:141-145(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95208357; PubMed=7900426;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
RN   [3]
RP   SEQUENCE OF 1-150 AND 1209-1392 FROM N.A.
RX   MEDLINE=87106859; PubMed=3542721;
RA   Fleig U.N., Pridmore R.D., Philippsen P.;
RT   "Construction of LYS2 cartridges for use in genetic manipulations of
RT   Saccharomyces cerevisiae.";
RL   Gene 46:237-245(1986).
RN   [4]
```

```
RP   SEQUENCE OF 1083-1392 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [5]
RP   SEQUENCE OF 1-130 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94039074; PubMed=7916691;
RA   Schaaff-Gerstenschlager I., Mannhaupt G., Vetter I., Zimmermann F.K.,
RA   Feldmann H.;
RT   "TKL2, a second transketolase gene of Saccharomyces cerevisiae.
RT   Cloning, sequence and deletion analysis of the gene.";
RL   Eur. J. Biochem. 217:487-492(1993).
CC   -!- FUNCTION: CATALYZES THE ACTIVATION OF ALPHA-AMINOADIPATE BY
CC       ATP-DEPENDENT ADENYLATION AND THE REDUCTION OF ACTIVATED
CC       ALPHA-AMINOADIPATE BY NADPH.
CC   -!- CATALYTIC ACTIVITY: L-2-aminoadipate 6-semialdehyde + NADP(+) +
CC       H(2)O = L-2-aminoadipate + NADPH.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE
CC       (POTENTIAL).
CC   -!- PATHWAY: Lysine biosynthesis; sixth step.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36287; AAA34747.1; -.
DR   EMBL; X66247; CAA46975.1; -.
DR   EMBL; X78993; CAA55617.1; -.
DR   EMBL; Z35984; CAA85072.1; -.
DR   EMBL; X73532; CAA51938.1; -.
DR   PIR; JU0448; YGBYAD.
DR   HSSP; P14687; 1AMU.
DR   SGD; S0000319; LYS2.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Lysine biosynthesis; Oxidoreductase; NADP; Phosphopantetheine.
FT   DOMAIN      848    917       ACYL CARRIER (ACP).
FT   BINDING     880    880       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1392 AA;  155345 MW;  F0083A80BC6F7FB5 CRC64;

  Query Match             18.6%;  Score 75.5;  DB 1;  Length 1392;
  Best Local Similarity   31.8%;  Pred. No. 2;
  Matches   21;  Conservative   12;  Mismatches   32;  Indels    1;  Gaps    1;

Qy         17 VRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPA 76
              |||   ::|    | :::|| ||  ||  |: | :    | |   : ||:  :| :||
Db        852 VRDLWLSIL-PTKPASVSPDDSFFDLGGHSILATKMIFTLKKKLQVDLPLGTIFKYPTIK 910

Qy         77 KLAVHL 82
                 |  :
Db        911 AFAAEI 916


RESULT 12
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
```

```
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match             17.6%;  Score 71.5;  DB 1;  Length 2505;
  Best Local Similarity   37.9%;  Pred. No. 10;
  Matches   25;  Conservative    7;  Mismatches   29;  Indels    5;  Gaps  2;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              :||      ||    :||:  || :|  |   | | |   || ||| || | :| :
Db       2111 VAHGDGEAQR----DLVK-AVAHILGIRDLAGINLDSSLADLGLDSLMGVEVRQILEREH 2165

Qy         61 GLRLPV 66
                | ||:
Db       2166 DLVLPI 2171
```

```
RESULT 13
ACP_STRCO
ID   ACP_STRCO      STANDARD;      PRT;    92 AA.
AC   Q9RK61;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acyl carrier protein (ACP).
GN   ACPP OR SCO0549 OR SCF11.29C.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- FUNCTION: Carrier of the growing fatty acid chain in fatty acid
CC       biosynthesis (By similarity).
CC   -!- PATHWAY: De novo fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- PTM: 4'-phosphopantetheine is transferred from CoA to a specific
CC       serine of apo-ACP by acpS. This modification is essential for
CC       activity because fatty acids are bound in thioester linkage to the
CC       sulfhydryl of the prosthetic group (By similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL132662; CAB59606.1; -.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
KW   Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;
KW   Complete proteome.
FT   BINDING      44     44       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   92 AA;  9876 MW;  8A1800D2395AA2B3 CRC64;

  Query Match             17.4%;  Score 71;  DB 1;  Length 92;
  Best Local Similarity   30.4%;  Pred. No. 0.33;
  Matches   24;  Conservative   15;  Mismatches   38;  Indels    2;  Gaps    2;

Qy          5 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFR-ALGFDSLTAVEFRNLLIKATGLR 63
              ::||:|  ||: :|| :|| :     : | || | :  || |||  :|    :     :
Db          3 STADER-QLLDELRDLLAATVEDLTVEEIGPDSSLQDDLGVDSLARLELVAAIEDRWQIE 61

Qy         64 LPVSLVFDHPTPAKLAVHL 82
              :|        |  ::| ||
Db         62 VPQEQADRLTTVRQIAAHL 80


RESULT 14
HTS1_COCCA
```

```
ID   HTS1_COCCA     STANDARD;      PRT;  5217 AA.
AC   Q01886;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   HC-toxin synthetase (EC 6.3.2.-) (HTS).
GN   HTS1.
OS   Cochliobolus carbonum (Bipolaris zeicola).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC   Pleosporales; Pleosporaceae; Cochliobolus.
OX   NCBI_TaxID=5017;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 90305 / SB111;
RX   MEDLINE=93100328; PubMed=1281482;
RA   Scott-Craig J.S., Panaccione D.G., Pocard J.-A., Walton J.D.;
RT   "The cyclic peptide synthetase catalyzing HC-toxin production in the
RT   filamentous fungus Cochliobolus carbonum is encoded by a
RT   15.7-kilobase open reading frame.";
RL   J. Biol. Chem. 267:26044-26049(1992).
RN   [2]
RP   FUNCTION.
RC   STRAIN=ATCC 90305 / SB111;
RX   MEDLINE=20138231; PubMed=10671527;
RA   Cheng Y.-Q., Walton J.D.;
RT   "A eukaryotic alanine racemase gene involved in cyclic peptide
RT   biosynthesis.";
RL   J. Biol. Chem. 275:4906-4911(2000).
CC   -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
CC       proline and AEO (2-amino-9,10-epoxi-8-oxodecanoic acid), and
CC       epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic
CC       tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
CC       it to D-Pro; also uses D-Ala as a substrate but this is epimerized
CC       from L-Ala by TOXG.
CC   -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M98024; AAA33023.1; -.
DR   HSSP; P14687; 1AMU.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF00668; Condensation; 5.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
KW   Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT   REPEAT      249    842       DOMAIN 1.
FT   REPEAT     1854   2452       DOMAIN 2.
FT   REPEAT     3006   3606       DOMAIN 3.
FT   REPEAT     4158   4738       DOMAIN 4.
FT   DOMAIN      762    840       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2384   2450       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3536   3604       ACYL CARRIER (ACP) 3.
FT   DOMAIN     4667   4736       ACYL CARRIER (ACP) 4.
FT   BINDING     803    803       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    2414   2414       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3568   3568       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    4700   4700       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   5217 AA;  574637 MW;  0331D9C5400163A5 CRC64;

  Query Match             17.3%;  Score 70.5;  DB 1;  Length 5217;
  Best Local Similarity   30.0%;  Pred. No. 29;
  Matches   18;  Conservative   13;  Mismatches   28;  Indels    1;  Gaps    1;

Qy         15 ELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPT 74
                :|:|     :|   |  |:    :|  :| ||: |::   || :  |: | |: :|  ||
Db       4669 KLLRQLWCKILPQLDESAVNKQDNFLGIGGDSIAAIKLVALL-RQHGISLAVAEIFTRPT 4727


RESULT 15
SRF1_BACSU
ID   SRF1_BACSU     STANDARD;      PRT;  3587 AA.
AC   P27206;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Surfactin synthetase subunit 1.
GN   SRFAA OR SRFA1 OR SRFA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA   Zuber P., Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
RL   Nucleic Acids Res. 21:93-97(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=93360813; PubMed=8355609;
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821-831(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [4]
RP   SEQUENCE OF 1-460 FROM N.A.
RX   MEDLINE=91154134; PubMed=1847909;
RA   Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA   Zuber P.;
RT   "srfA is an operon required for surfactin production, competence
RT   development, and efficient sporulation in Bacillus subtilis.";
RL   J. Bacteriol. 173:1770-1778(1991).
RN   [5]
RP   SEQUENCE OF 1-38 FROM N.A.
RX   MEDLINE=91358326; PubMed=1715856;
RA   Nakano M.M., Xia L., Zuber P.;
RT   "Transcription initiation region of the srfA operon, which is
RT   controlled by the comP-comA signal transduction system in Bacillus
RT   subtilis.";
RL   J. Bacteriol. 173:5487-5493(1991).
RN   [6]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95219080; PubMed=7704255;
RA   Fujishima Y., Yamane K.;
RT   "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
```

```
RT   of srfAA of the Bacillus subtilis chromosome.";
RL   Microbiology 141:277-279(1995).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13262; BAA02522.1; -.
DR   EMBL; X70356; CAA49816.1; -.
DR   EMBL; D50453; BAA08982.1; -.
DR   EMBL; M59939; AAA22815.1; -.
DR   EMBL; M64702; AAA22816.1; -.
DR   EMBL; D30762; BAA21034.1; -.
DR   EMBL; Z99105; CAB12142.1; -.
DR   PIR; S35517; S35517.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10168; srfAA.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW   Multifunctional enzyme; Repeat; Complete proteome.
FT   REPEAT        ?   1047       DOMAIN 1 (GLU-ACTIVATING).
FT   REPEAT        ?   2086       DOMAIN 2 (LEU-ACTIVATING).
FT   REPEAT        ?   3114       DOMAIN 3 (D-LEU-ACTIVATING).
FT   DOMAIN      976   1043       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2015   2082       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3043   3109       ACYL CARRIER (ACP) 3.
FT   BINDING    1006   1006       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2045   2045       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3073   3073       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    146    146       I -> M (IN REF. 2 AND 3).
FT   CONFLICT    151    151       L -> V (IN REF. 2 AND 3).
FT   CONFLICT    165    165       A -> G (IN REF. 2, 3 AND 4).
FT   CONFLICT    281    281       T -> Q (IN REF. 2 AND 3).
FT   CONFLICT    460    460       D -> T (IN REF. 3 AND 3).
FT   CONFLICT    540    540       P -> A (IN REF. 2 AND 3).
FT   CONFLICT    562    562       I -> Y (IN REF. 2 AND 3).
FT   CONFLICT    639    640       GS -> PT (IN REF. 2 AND 3).
FT   CONFLICT    644    644       R -> P (IN REF. 2 AND 3).
FT   CONFLICT    647    647       L -> F (IN REF. 2 AND 3).
FT   CONFLICT    649    649       A -> P (IN REF. 2 AND 3).
FT   CONFLICT   1026   1026       L -> Q (IN REF. 2 AND 3).
FT   CONFLICT   1065   1116       SLAQQRTYIVSQFEDAGVGYNMPAAAILEGPLDIQKLERAF
FT                                QGLIRRHESLR -> HWHSSEHISSASSRMRESAIHASSSN
FT                                SEGFRYSKAGARISGINPTPRVIE (IN REF. 2 AND
FT                                3).
FT   CONFLICT   1132   1133       DS -> VC (IN REF. 2 AND 3).
FT   CONFLICT   1164   1164       R -> A (IN REF. 2 AND 3).
FT   CONFLICT   1458   1458       V -> L (IN REF. 2 AND 3).
FT   CONFLICT   1850   1850       S -> T (IN REF. 2 AND 3).
FT   CONFLICT   1894   1894       D -> T (IN REF. 2 AND 3).
FT   CONFLICT   1911   1911       R -> A (IN REF. 2 AND 3).
FT   CONFLICT   1973   1980       LRCLSKWT -> PAVFIQMD (IN REF. 2 AND 3).
FT   CONFLICT   2052   2052       L -> P (IN REF. 2 AND 3).
FT   CONFLICT   2218   2218       H -> Q (IN REF. 2 AND 3).
FT   CONFLICT   2265   2266       QQ -> HR (IN REF. 2 AND 3).
FT   CONFLICT   2291   2291       V -> L (IN REF. 2 AND 3).
FT   CONFLICT   2349   2349       S -> C (IN REF. 2 AND 3).
FT   CONFLICT   2428   2428       R -> S (IN REF. 2 AND 3).
FT   CONFLICT   2611   2612       RC -> AV (IN REF. 2 AND 3).
FT   CONFLICT   2685   2685       E -> ENPE (IN REF. 2 AND 3).
FT   CONFLICT   2755   2755       T -> S (IN REF. 2 AND 3).
FT   CONFLICT   2896   2897       SP -> TA (IN REF. 2 AND 3).
FT   CONFLICT   3025   3025       P -> N (IN REF. 2 AND 3).
FT   CONFLICT   3096   3096       F -> N (IN REF. 2 AND 3).
FT   CONFLICT   3271   3271       A -> S (IN REF. 2 AND 3).
FT   CONFLICT   3316   3316       R -> S (IN REF. 2 AND 3).
FT   CONFLICT   3451   3451       Y -> S (IN REF. 2 AND 3).
FT   CONFLICT   3483   3484       DE -> HQ (IN REF. 2 AND 3).
FT   CONFLICT   3487   3490       DAGL -> HPPF (IN REF. 2 AND 3).
FT   CONFLICT   3493   3493       R -> T (IN REF. 2 AND 3).
FT   CONFLICT   3495   3495       E -> H (IN REF. 2 AND 3).
FT   CONFLICT   3499   3500       GQ -> PH (IN REF. 2 AND 3).
FT   CONFLICT   3507   3507       R -> E (IN REF. 2 AND 3).
SQ   SEQUENCE   3587 AA;  402424 MW;  A1E2DABFF93EDE3A CRC64;
```

```
  Query Match           17.2%;  Score 70;  DB 1;  Length 3587;
  Best Local Similarity  35.7%;  Pred. No. 22;
  Matches   15;  Conservative    8;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         38 SFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLA 79
              :|   |  || |:     : | ||: :|:  :|:|||   ||
Db        998 NFFETGGHSLKAMTLLTKIHKETGIEIPLQFLFEHPTITALA 1039


Search completed: June 17, 2003, 13:02:36
Job time : 1.89424 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 1.83105 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_1979_2060
Perfect score: 407
Sequence: 1 LAHQTSADQRAALLELVRDH.....RLPVSLVFDHPTPAKLAVHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	229	56.3	3739	2	T17410	polyketide synthas
2	225	55.3	1562	2	T17411	polyketide synthas
3	224	55.0	1763	2	T17465	rifamycin polyketi
4	223	54.8	2100	2	T03223	probable polyketid
5	221	54.3	5069	2	T17464	rifamycin polyketi
6	221	54.3	6420	2	T30283	polyketide synthas
7	216.5	53.2	1762	2	T03222	probable polyketid
8	216	53.1	3413	2	T17467	rifamycin polyketi
9	216	53.1	4151	2	G70944	probable polyketid
10	215	52.8	8563	2	T30226	polyketide synthas
11	214.5	52.7	10223	2	T30225	polyketide synthas
12	212	52.1	1728	2	T17466	rifamycin polyketi
13	212	52.1	6260	2	T30228	polyketide synthas
14	205	50.4	3172	2	S22012	erythronolide synt
15	205	50.4	3178	2	S13595	6-deoxyerythronoli
16	203	49.9	1937	2	T03224	probable polyketid
17	202.5	49.8	7576	2	T17428	FK506 polyketide s
18	202	49.6	3491	2	T43231	probable 6-deoxyer
19	199	48.9	4735	2	T17463	rifamycin polyketi
20	197	48.4	4613	2	T17409	polyketide synthas
21	196	48.2	1198	2	T28678	polyketide synthas
22	192.5	47.3	2723	2	T03221	probable polyketid
23	186	45.7	3519	2	S43048	polyketide synthas
24	180	44.2	3573	2	S23070	erythronolide synt
25	176	43.2	1346	2	T17412	polyketide synthas
26	172.5	42.4	1616	2	G70668	polyketide synthas
27	172	42.3	502	2	A70985	probable polyketid
28	169	41.5	1402	2	D70634	probable polyketid
29	169	41.5	2103	2	G86925	probable polyketid
30	167.5	41.2	2126	2	H70621	probable polyketid
31	151	37.1	2124	2	T28658	polyketide synthas
32	139	34.2	2201	2	S73014	polyketide synthas
33	135.5	33.3	1017	2	B70985	probable polyketid
34	122	30.0	2188	2	A70984	probable polyketid
35	117	28.7	8243	2	T31307	type I fatty acid
36	113	27.8	2478	2	AH2140	polyketide synthas
37	104.5	25.7	4427	2	PN0637	polyketide synthas
38	103.5	25.4	4930	2	E69679	polyketide synthet
39	101.5	24.9	7829	2	T15789	hypothetical prote
40	99.5	24.4	2458	2	T17420	probable polyketid
41	99	24.3	1540	2	H87203	polyketide synthas
42	98	24.1	1538	2	E70874	probable ppsB prot
43	97.5	24.0	958	2	S73012	polyketide synthas
44	97.5	24.0	1570	2	AC2012	hypothetical prote
45	97.5	24.0	1871	2	A87204	polyketide synthas

ALIGNMENTS

RESULT 1
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

Query Match 56.3%; Score 229; DB 2; Length 3739;
Best Local Similarity 59.8%; Pred. No. 8.9e-17;
Matches 49; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641
```

RESULT 2
T17411
polyketide synthase III - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17411
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17411
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1562 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C;Genetics:
A;Gene: pikAIII

```
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1403-1474/Domain: acyl carrier protein homology <ACP>

  Query Match             55.3%;  Score 225;  DB 2;  Length 1562;
  Best Local Similarity   60.0%;  Pred. No. 9.4e-17;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |     | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
                |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 3
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>

  Query Match             55.0%;  Score 224;  DB 2;  Length 1763;
  Best Local Similarity   57.3%;  Pred. No. 1.4e-16;
  Matches   47;  Conservative    9;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  | |:| | ||:|||  || || |||   :  | :|:  ||||||:|| || | : |
Db       1605 LAGLTVAEQEALLLDLVRGQVAVVLGHADSSGVRADAAFKDAGFDSLTSVELRNRLRETT 1664

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||:|| :|||||| |  || ||
Db       1665 GLKLPATLVFDHPNPLALARHL 1686


RESULT 4
T03223
probable polyketide synthase module 3 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03223
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2100 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38063.1; PID:g2624950
A;Experimental source: ATCC 29253
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;566-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1952-2027/Domain: acyl carrier protein homology <ACP>

  Query Match             54.8%;  Score 223;  DB 2;  Length 2100;
  Best Local Similarity   58.7%;  Pred. No. 2.2e-16;
  Matches   44;  Conservative   12;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::|| :|||||   | || |:  : ::| |:|:  |||||||||||| :  ||||||| :
Db       1949 ERRAVVLELVRAQAATVLGHSGAEHVSPTQAFKDSGFDSLTAVEFRNRVTTATGLRLPAT 2008

Qy         68 LVFDHPTPAKLAVHL 82
              |:|||||  || ||
Db       2009 LIFDHPTPNTLADHL 2023


RESULT 5
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             54.3%;  Score 221;  DB 2;  Length 5069;
  Best Local Similarity   58.5%;  Pred. No. 9.8e-16;
  Matches   48;  Conservative    8;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    |||   ||:|||  ||||| |||  |:  | :|:  |||||||||| ||  :  ||
Db       3228 LAGLAPADQLTLLLDLVRAQVAAVLGHADASAVRVDTAFKDAGFDSLTAVELRNRMRTAT 3287

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||:|| :||||:| |  || ||
Db       3288 GLKLPATLVFDYPNPQALARHL 3309


RESULT 6
T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30283
R;Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A;Title: Structural organization of a multifunctional polyketide synthase involved in t
A;Reference number: Z20806; MUID:97217427; PMID:9063448
A;Accession: T30283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6420 <MOT>
A;Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C;Genetics:
A;Note: fkbA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
C;Keywords: carrier protein
F;51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3391-3462/Domain: acyl carrier protein homology <ACP1>
F;3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
```

```
F;5307-5378/Domain: acyl carrier protein homology <ACP2>
F;5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6293-6364/Domain: acyl carrier protein homology <ACP3>

  Query Match             54.3%;  Score 221;  DB 2;  Length 6420;
  Best Local Similarity   57.3%;  Pred. No. 1.3e-15;
  Matches   47;  Conservative    9;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    :||:  ||: || |  |||| |||   |||  :|: || |||||:| || | :||
Db       3381 LAELPAADRDTALVTLVSDATAAVLGHADASGIAPTTTFKDLGIDSLTAIELRNRLAEAT 3440

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||  ::|||||||  ||  |
Db       3441 GLRLSATMVFDHPTPRVLAAEL 3462


RESULT 7
T03222
probable polyketide synthase module 2 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Nov-2000
C;Accession: T03222
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03222
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1762 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38062.1; PID:g2624949
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;550-822/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1598-1669/Domain: acyl carrier protein homology <ACP>

  Query Match             53.2%;  Score 216.5;  DB 2;  Length 1762;
  Best Local Similarity   56.5%;  Pred. No. 9.6e-16;
  Matches   48;  Conservative    8;  Mismatches   20;  Indels    9;  Gaps    1;

Qy          7 ADQRAAL---------LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLI 57
              ||: |||         |:|||    | || |:|    | | |:|:  ||||||||||| :
Db       1585 ADRIAALMPEERDRAVLDLVRQQSARVLGHSDADRIDPGQAFKDTGFDSLTAVEFRNRVT 1644

Qy         58 KATGLRLPVSLVFDHPTPAKLAVHL 82
              ||||||| :|:||||||  || ||
Db       1645 TATGLRLPATLIFDHPTPNALADHL 1669


RESULT 8
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

  Query Match             53.1%;  Score 216;  DB 2;  Length 3413;
  Best Local Similarity   54.9%;  Pred. No. 2.3e-15;
  Matches   45;  Conservative   14;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    |:| | ||::||  || || || |:|:  | :|:  ||||||:|| || | :|:
Db       1598 LAGLAPAEQEALLLDVVRTQVALVLGHAGPEAVRADTAFKDTGFDSLTSVELRNRLREAS 1657

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||:|| :||||:|||  || :|
Db       1658 GLKLPATLVFDYPTPVALARYL 1679


RESULT 9
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g28967
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoac
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

  Query Match             53.1%;  Score 216;  DB 2;  Length 4151;
  Best Local Similarity   54.7%;  Pred. No. 2.8e-15;
  Matches   47;  Conservative   14;  Mismatches   21;  Indels    4;  Gaps    1;

Qy          1 LAHQ----TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLL 56
              |||:       |:| | || ||| |:| || :  |:|| ||::|: ||||||||| || |
Db       1949 LAHRLHGLPEAEQHAVLLGLVRLHIATVLGNITPEAIDPDKAFQDLGFDSLTAVEMRNRL 2008

Qy         57 IKATGLRLPVSLVFDHPTPAKLAVHL 82
                |||| |  :|:||:||| :|| ::
Db       2009 KSATGLSLSPTLIFDYPTPNRLASYI 2034


RESULT 10
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Stau
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             52.8%;  Score 215;  DB 2;  Length 8563;
  Best Local Similarity   58.5%;  Pred. No. 8.2e-15;
  Matches   48;  Conservative    7;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    |::  |||:|| |  | || |||   ::    || || ||||||| || | |||
Db       4777 LAALAPAEREKALLKLVCDSAATVLGHADTSTVSVAAVFRDLGVDSLTAVELRNSLAKAT 4836

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :||||:|||  ||| |
Db       4837 GLRLPATLVFDYPTPTALAVRL 4858


RESULT 11
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match             52.7%;  Score 214.5;  DB 2;  Length 10223;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-14;
  Matches   46;  Conservative   11;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          8 DQRA-ALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPV 66
              ::|| |||::|||  | || ||| ::|    :|: || ||||||| || | ||||||||
Db      10065 EERAKALLKVVRDSAATVLGHADARSIPATGAFKDLGVDSLTAVELRNSLTKATGLRLPA 10124

Qy         67 SLVFDHPTPAKLAVHL 82
              ::|||:|||| ||  |
Db      10125 TMVFDYPTPADLAARL 10140


RESULT 12
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17466
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17466
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1728 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1573-1644/Domain: acyl carrier protein homology <ACP>

  Query Match             52.1%;  Score 212;  DB 2;  Length 1728;
  Best Local Similarity   53.7%;  Pred. No. 3e-15;
  Matches   44;  Conservative   14;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    : :| | |::|||  ||||| || | |:  | :|:  ||||||:|: || | ::|
Db       1563 LAGLGAPEQEALLVDLVRGQVAAVLGHAGPDAVRADTAFKDAGFDSLTSVDLRNRLREST 1622

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||:|| :| ||:|||  || ||
Db       1623 GLKLPATLAFDYPTPLVLARHL 1644


RESULT 13
T30228
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30228
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30228
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6260 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987102; PIDN:CAA60462.1
C;Genetics:
A;Gene: rapC
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
C;Keywords: carrier protein
F;55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1536-1607/Domain: acyl carrier protein homology <ACP1>
F;1651-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3123-3194/Domain: acyl carrier protein homology <ACP2>
F;3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3725-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5114-5185/Domain: acyl carrier protein homology <ACP3>
F;5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6135-6206/Domain: acyl carrier protein homology <ACP4>

  Query Match             52.1%;  Score 212;  DB 2;  Length 6260;
  Best Local Similarity   58.5%;  Pred. No. 1.3e-14;
```

```
  Matches   48;  Conservative    7;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||    |::  |||:||||  | || |||   |    :|: || ||||||| || | |||
Db       3113 LAALAPAEREDALLKLVRDSAALVLGHADASTIPAAAAFKDLGIDSLTAVELRNSLAKAT 3172

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| : |||:|||| ||  |
Db       3173 GLRLPNTTVFDYPTPAILATRL 3194


RESULT 14
S22012
erythronolide synthase (EC 2.3.1.94) III - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase III; protein A
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S22012; S34784; S23206
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22012
A;Molecule type: DNA
A;Residues: 1-3172 <BEV1>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44449.1; PID:g46979
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S34784
A;Molecule type: DNA
A;Residues: 1-715,717-728,'C',729-886,888-894,'QR',895-910,'E',912-1276,'RR',1279-1284,'
EV2>
A;Cross-references: EMBL:X62569
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23206
A;Molecule type: protein
A;Residues: 2-11,'XXX',15 <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;555-835/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1117-1297/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1392-1463/Domain: acyl carrier protein homology <ACP1>
F;1511-1914/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2022-2300/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2557-2734/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2819-2890/Domain: acyl carrier protein homology <ACP2>
F;2959-3153/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

  Query Match             50.4%;  Score 205;  DB 2;  Length 3172;
  Best Local Similarity   58.6%;  Pred. No. 3.6e-14;
  Matches   41;  Conservative    8;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         13 LLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDH 72
              |||    ||||:| |: | |:  || |  |||||||||  || | :|||| || :|||:|
Db       2821 LLEFTHSHVAAILGHSSPDAVGQDQPFTELGFDSLTAVGLRNQLQQATGLALPATLVFEH 2880

Qy         73 PTPAKLAVHL 82
              ||  :|| |:
Db       2881 PTVRRLADHI 2890


RESULT 15
S13595
6-deoxyerythronolide B synthase - Saccharopolyspora erythraea
N;Alternate names: protein A
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S13595
R;Cortes, J.; Haydock, S.F.; Roberts, G.A.; Bevitt, D.J.; Leadlay, P.F.
Nature 348, 176-178, 1990
A;Title: An unusually large multifunctional polypeptide in the erythromycin-producing p
A;Reference number: S13595; MUID:91043075; PMID:2234082
A;Accession: S13595
A;Molecule type: DNA
A;Residues: 1-3178 <COR>
A;Cross-references: EMBL:X56107; NID:g46975; PIDN:CAA39583.1; PID:g46976
A;Experimental source: strain NRRL 2338
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
-carrier-protein] S-malonyltransferase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; multifunctional
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;553-833/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1123-1303/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1398-1469/Domain: acyl carrier protein homology <ACP1>
F;1517-1920/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2028-2306/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2563-2740/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2825-2896/Domain: acyl carrier protein homology <ACP2>
F;2965-3159/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

  Query Match             50.4%;  Score 205;  DB 2;  Length 3178;
  Best Local Similarity   58.6%;  Pred. No. 3.6e-14;
  Matches   41;  Conservative    8;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         13 LLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDH 72
              |||    ||||:| |: | |:  || |  |||||||||  || | :|||| || :|||:|
Db       2827 LLEFTHSHVAAILGHSSPDAVGQDQPFTELGFDSLTAVGLRNQLQQATGLALPATLVFEH 2886

Qy         73 PTPAKLAVHL 82
              ||  :|| |:
Db       2887 PTVRRLADHI 2896


Search completed: June 17, 2003, 13:12:22
Job time : 2.83105 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 2.55496 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_1979_2060
Perfect score: 407
Sequence: 1 LAHQTSADQRAALLELVRDH.....RLPVSLVFDHPTPAKLAVHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	229	56.3	3739	9	US-09-860-846-33	Sequence 33, Appl
2	229	56.3	3739	9	US-09-988-384B-33	Sequence 33, Appl
3	229	56.3	3739	9	US-09-836-821-33	Sequence 33, Appl
4	229	56.3	3739	9	US-09-793-708-2	Sequence 2, Appli
5	229	56.3	3739	10	US-09-861-289-33	Sequence 33, Appl
6	229	56.3	11877	9	US-09-860-846-6	Sequence 6, Appli
7	229	56.3	11877	9	US-09-836-821-6	Sequence 6, Appli
8	229	56.3	11877	10	US-09-861-289-6	Sequence 6, Appli
9	229	56.3	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	225	55.3	1562	9	US-09-860-846-35	Sequence 35, Appl
11	225	55.3	1562	9	US-09-988-384B-35	Sequence 35, Appl
12	225	55.3	1562	9	US-09-836-821-35	Sequence 35, Appl
13	225	55.3	1562	9	US-09-793-708-3	Sequence 3, Appli
14	225	55.3	1562	10	US-09-861-289-35	Sequence 35, Appl
15	224	55.0	3816	9	US-09-808-880-3	Sequence 3, Appli
16	223	54.8	4150	9	US-09-808-880-2	Sequence 2, Appli
17	211	51.8	5215	9	US-09-860-846-2	Sequence 2, Appli
18	211	51.8	5215	9	US-09-988-384B-2	Sequence 2, Appli
19	211	51.8	5215	9	US-09-836-821-2	Sequence 2, Appli
20	211	51.8	5215	10	US-09-861-289-2	Sequence 2, Appli
21	197	48.4	4551	9	US-09-793-708-1	Sequence 1, Appli
22	197	48.4	4613	9	US-09-860-846-31	Sequence 31, Appl
23	197	48.4	4613	9	US-09-988-384B-31	Sequence 31, Appl
24	197	48.4	4613	9	US-09-836-821-31	Sequence 31, Appl
25	197	48.4	4613	10	US-09-861-289-31	Sequence 31, Appl
26	186	45.7	3519	9	US-09-808-880-4	Sequence 4, Appli
27	178	43.7	1346	9	US-09-793-708-4	Sequence 4, Appli
28	176	43.2	1346	9	US-09-860-846-37	Sequence 37, Appl
29	176	43.2	1346	9	US-09-988-384B-37	Sequence 37, Appl
30	176	43.2	1346	9	US-09-836-821-37	Sequence 37, Appl
31	176	43.2	1346	10	US-09-861-289-37	Sequence 37, Appl
32	172.5	42.4	1616	9	US-09-712-363-262	Sequence 262, App
33	169	41.5	1402	9	US-09-712-363-166	Sequence 166, App
34	163	40.0	3798	9	US-10-014-717-6	Sequence 6, Appli
35	163	40.0	7257	9	US-10-014-717-5	Sequence 5, Appli
36	112	27.5	1832	9	US-10-014-717-4	Sequence 4, Appli
37	101	24.8	352	10	US-09-073-009-14	Sequence 14, Appl
38	101	24.8	352	10	US-09-023-588-14	Sequence 14, Appl
39	101	24.8	352	10	US-09-793-306-14	Sequence 14, Appl
40	98.5	24.2	2439	9	US-10-014-717-7	Sequence 7, Appli
41	97.5	24.0	1421	9	US-10-014-717-2	Sequence 2, Appli
42	86	21.1	1827	9	US-09-712-363-261	Sequence 261, App
43	85.5	21.0	1812	10	US-09-775-938A-38	Sequence 38, Appl
44	78.5	19.3	4999	9	US-09-976-059-15	Sequence 15, Appl
45	77	18.9	1051	9	US-09-976-059-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33

Query Match 56.3%; Score 229; DB 9; Length 3739;
Best Local Similarity 59.8%; Pred. No. 1.1e-18;
Matches 49; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  ||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
                ||||| :||||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641
```

RESULT 2
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             56.3%;  Score 229;  DB 9;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 1.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 3
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             56.3%;  Score 229;  DB 9;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 1.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 4
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-2

  Query Match             56.3%;  Score 229;  DB 9;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 1.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 5
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             56.3%;  Score 229;  DB 10;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 1.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
```

```
                  ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641
```

RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

```
  Query Match           56.3%;  Score 229;  DB 9;  Length 11877;
  Best Local Similarity   59.8%;  Pred. No. 5.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590
```

RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-6

```
  Query Match           56.3%;  Score 229;  DB 9;  Length 11877;
  Best Local Similarity   59.8%;  Pred. No. 5.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
```

```
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590
```

RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6

```
  Query Match           56.3%;  Score 229;  DB 10;  Length 11877;
  Best Local Similarity   59.8%;  Pred. No. 5.1e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590
```

RESULT 9
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
; LENGTH: 12199
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

```
  Query Match           56.3%;  Score 229;  DB 9;  Length 12199;
  Best Local Similarity   59.8%;  Pred. No. 5.3e-18;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       8831 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8890

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
```

```
Db       8891 GLRLPATLVFDHPTPGELAGHL 8912


RESULT 10
US-09-860-846-35
; Sequence 35, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-35

  Query Match             55.3%;  Score 225;  DB 9;  Length 1562;
  Best Local Similarity   60.0%;  Pred. No. 1.1e-18;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |     | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
                |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 11
US-09-988-384B-35
; Sequence 35, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-35

  Query Match             55.3%;  Score 225;  DB 9;  Length 1562;
  Best Local Similarity   60.0%;  Pred. No. 1.1e-18;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |     | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
                |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 12
US-09-836-821-35
; Sequence 35, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-35

  Query Match             55.3%;  Score 225;  DB 9;  Length 1562;
  Best Local Similarity   60.0%;  Pred. No. 1.1e-18;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |     | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
                |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 13
US-09-793-708-3
; Sequence 3, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-3
```

```
  Query Match           55.3%;  Score 225;  DB 9;  Length 1562;
  Best Local Similarity  60.0%;  Pred. No. 1.1e-18;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |     | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
               |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 14
US-09-861-289-35
; Sequence 35, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-35

  Query Match           55.3%;  Score 225;  DB 10;  Length 1562;
  Best Local Similarity  60.0%;  Pred. No. 1.1e-18;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |     | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
               |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 15
US-09-808-880-3
; Sequence 3, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

  Query Match           55.0%;  Score 224;  DB 9;  Length 3816;
  Best Local Similarity  62.7%;  Pred. No. 4.7e-18;
  Matches   47;  Conservative    9;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              | | || :||| ||| || |: |:||  :|:|| :||||||||| || |    |||||| :
Db       3632 DPRRALRDLVRGHVALVLGHSGPEAIDAEQAFRDIGFDSLTAVELRNRLNAETGLRLPGT 3691

Qy         68 LVFDHPTPAKLAVHL 82
               ||||:| |: || ||
Db       3692 LVFDYPNPSALADHL 3706


Search completed: June 17, 2003, 13:46:43
Job time : 2.55496 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 1.49039 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_1979_2060
Perfect score: 407
Sequence: 1 LAHQTSADQRAALLELVRDH.....RLPVSLVFDHPTPAKLAVHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	229	56.3	3739	3	US-09-320-878-2	Sequence 2, Appli
2	229	56.3	3739	4	US-09-105-537-33	Sequence 33, Appl
3	229	56.3	11877	4	US-09-105-537-6	Sequence 6, Appli
4	225	55.3	1562	3	US-09-320-878-3	Sequence 3, Appli
5	225	55.3	1562	4	US-09-105-537-35	Sequence 35, Appl
6	224	55.0	3816	4	US-09-428-517-3	Sequence 3, Appli
7	223	54.8	3724	2	US-08-804-227C-10	Sequence 10, Appl
8	223	54.8	3724	2	US-08-804-198-4	Sequence 4, Appli
9	223	54.8	4150	4	US-09-428-517-2	Sequence 2, Appli
10	222	54.5	4545	2	US-08-804-227C-14	Sequence 14, Appl
11	222	54.5	4550	2	US-08-804-227C-8	Sequence 8, Appli
12	222	54.5	4550	2	US-08-804-198-2	Sequence 2, Appli
13	219	53.8	1996	2	US-08-804-227C-9	Sequence 9, Appli
14	219	53.8	1996	2	US-08-804-198-3	Sequence 3, Appli
15	219	53.8	3729	2	US-08-804-227C-4	Sequence 4, Appli
16	212	52.1	2152	4	US-09-036-987A-3	Sequence 3, Appli
17	212	52.1	2152	4	US-09-370-700-3	Sequence 3, Appli
18	211	51.8	5215	4	US-09-105-537-2	Sequence 2, Appli
19	210	51.6	3170	4	US-09-036-987A-4	Sequence 4, Appli
20	210	51.6	3170	4	US-09-370-700-4	Sequence 4, Appli
21	209.5	51.5	185	4	US-09-154-083-14	Sequence 14, Appl
22	205	50.4	3170	2	US-07-642-734C-5	Sequence 5, Appli
23	205	50.4	3170	3	US-08-439-009A-5	Sequence 5, Appli
24	205	50.4	3567	2	US-07-642-734C-4	Sequence 4, Appli
25	205	50.4	3567	3	US-08-439-009A-4	Sequence 4, Appli
26	202	49.6	3491	2	US-07-642-734C-2	Sequence 2, Appli
27	202	49.6	3491	3	US-08-439-009A-2	Sequence 2, Appli
28	197	48.4	4551	3	US-09-320-878-1	Sequence 1, Appli
29	197	48.4	4613	4	US-09-105-537-31	Sequence 31, Appl
30	195	47.9	4928	4	US-09-036-987A-5	Sequence 5, Appli
31	195	47.9	4928	4	US-09-370-700-5	Sequence 5, Appli
32	191	46.9	5588	4	US-09-036-987A-6	Sequence 6, Appli
33	191	46.9	5588	4	US-09-370-700-6	Sequence 6, Appli
34	186	45.7	1864	2	US-08-804-227C-3	Sequence 3, Appli
35	186	45.7	3519	4	US-09-428-517-4	Sequence 4, Appli
36	183	45.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
37	182	44.7	502	4	US-09-413-814-106	Sequence 106, App
38	178	43.7	1346	3	US-09-320-878-4	Sequence 4, Appli
39	178	43.7	2595	4	US-09-036-987A-2	Sequence 2, Appli
40	178	43.7	2595	4	US-09-370-700-2	Sequence 2, Appli
41	176	43.2	1346	4	US-09-105-537-37	Sequence 37, Appl
42	176	43.2	1611	2	US-08-804-227C-5	Sequence 5, Appli
43	174	42.8	1841	2	US-08-804-227C-6	Sequence 6, Appli
44	165	40.5	1891	2	US-08-804-227C-12	Sequence 12, Appl
45	165	40.5	1891	2	US-08-804-198-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2

Query Match 56.3%; Score 229; DB 3; Length 3739;
Best Local Similarity 59.8%; Pred. No. 9.8e-22;
Matches 49; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641
```

RESULT 2
US-09-105-537-33
; Sequence 33, Application US/09105537A

```
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-33

  Query Match           56.3%;  Score 229;  DB 4;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 9.8e-22;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || | |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 3
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match           56.3%;  Score 229;  DB 4;  Length 11877;
  Best Local Similarity   59.8%;  Pred. No. 5e-21;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || | |  :  :::||  |||||||||| || |  ||
Db       8509 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8568

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       8569 GLRLPATLVFDHPTPGELAGHL 8590


RESULT 4
US-09-320-878-3
; Sequence 3, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-3

  Query Match           55.3%;  Score 225;  DB 3;  Length 1562;
  Best Local Similarity   60.0%;  Pred. No. 1e-21;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |    | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
              |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474


RESULT 5
US-09-105-537-35
; Sequence 35, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
;   LENGTH: 1562
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-35

  Query Match           55.3%;  Score 225;  DB 4;  Length 1562;
  Best Local Similarity   60.0%;  Pred. No. 1e-21;
  Matches   45;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |    | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
              |||||||| || ||
Db       1460 TVFDHPTPAALAAHL 1474
```

```
RESULT 6
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3

  Query Match             55.0%;  Score 224;  DB 4;  Length 3816;
  Best Local Similarity   62.7%;  Pred. No. 4.9e-21;
  Matches   47;  Conservative    9;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              | | || :||| ||| || |: |:||  :|:|| :||||||||| || |   ||||||  :
Db       3632 DPRRALRDLVRGHVALVLGHSGPEAIDAEQAFRDIGFDSLTAVELRNRLNAETGLRLPGT 3691

Qy         68 LVFDHPTPAKLAVHL 82
              ||||:| |: || ||
Db       3692 LVFDYPNPSALADHL 3706


RESULT 7
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-10

  Query Match             54.8%;  Score 223;  DB 2;  Length 3724;
  Best Local Similarity   54.9%;  Pred. No. 6.4e-21;
  Matches   45;  Conservative   12;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  :: ::   || || :|||||| |     : ||: || :|||||||| || :   |
Db       3548 LAGLSAEERHDRLLGLVGEHVAAVLGHGSAAEVRPDRPFREVGFDSLTAVELRNRMAAVT 3607

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              |:||| :||||||||| |: ||
Db       3608 GVRLPATLVFDHPTPAALSSHL 3629


RESULT 8
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-4

  Query Match             54.8%;  Score 223;  DB 2;  Length 3724;
  Best Local Similarity   54.9%;  Pred. No. 6.4e-21;
  Matches   45;  Conservative   12;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  :: ::   || || :|||||| |     : ||: || :|||||||| || :   |
Db       3548 LAGLSAEERHDRLLGLVGEHVAAVLGHGSAAEVRPDRPFREVGFDSLTAVELRNRMAAVT 3607

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
```

```
              |:||| :|||||||||| |: ||
Db       3608 GVRLPATLVFDHPTPAALSSHL 3629


RESULT 9
US-09-428-517-2
; Sequence 2, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-2

  Query Match             54.8%;  Score 223;  DB 4;  Length 4150;
  Best Local Similarity   56.1%;  Pred. No. 7.5e-21;
  Matches   46;  Conservative   14;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |: :|   |::|||:  | || |:  :|:|||::|: ||||||:|| || |  ||
Db       2441 LAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVAPDRAFKDLGFDSLTSVELRNRLNTAT 2500

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||||: |||:  || || ||
Db       2501 GLRLPVTAVFDYARPAALAGHL 2522


RESULT 10
US-08-804-227C-14
; Sequence 14, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-14

  Query Match             54.5%;  Score 222;  DB 2;  Length 4545;
  Best Local Similarity   57.9%;  Pred. No. 1.2e-20;
  Matches   44;  Conservative   10;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          7 ADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPV 66
              |::| ||| |||   | :||||   |: |: :||: |||||| :| || |  |||| ||
Db       2534 AERRRALLTLVRSEAAGILRHASADAVDPELAFRSAGFDSLTVLELRNRLTAATGLNLPN 2593

Qy         67 SLVFDHPTPAKLAVHL 82
              :|:||||||  || ||
Db       2594 TLLFDHPTPLSLASHL 2609


RESULT 11
US-08-804-227C-8
; Sequence 8, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4550 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-8

  Query Match             54.5%;  Score 222;  DB 2;  Length 4550;
  Best Local Similarity   57.9%;  Pred. No. 1.2e-20;
  Matches   44;  Conservative   10;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          7 ADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPV 66
              |::| ||| |||   | :||||   |: |: :||: |||||| :| || |  |||| ||
```

```
Db       2539 AERRRALLTLVRSEAAGILRHASADAVDPELAFRSAGFDSLTVLELRNRLTAATGLNLPN 2598

Qy         67 SLVFDHPTPAKLAVHL 82
              :|:||||||  || ||
Db       2599 TLLFDHPTPLSLASHL 2614


RESULT 12
US-08-804-198-2
; Sequence 2, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4550 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-2

  Query Match             54.5%;  Score 222;  DB 2;  Length 4550;
  Best Local Similarity   57.9%;  Pred. No. 1.2e-20;
  Matches   44;  Conservative   10;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          7 ADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPV 66
              |::| ||| |||   | :||||   |: |: :||: |||||| :| || |  |||| ||
Db       2539 AERRRALLTLVRSEAAGILRHASADAVDPELAFRSAGFDSLTVLELRNRLTAATGLNLPN 2598

Qy         67 SLVFDHPTPAKLAVHL 82
              :|:||||||  || ||
Db       2599 TLLFDHPTPLSLASHL 2614


RESULT 13
US-08-804-227C-9
; Sequence 9, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1996 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-9

  Query Match             53.8%;  Score 219;  DB 2;  Length 1996;
  Best Local Similarity   58.2%;  Pred. No. 9.4e-21;
  Matches   46;  Conservative   11;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || ::: :: | ||:|||  ||||| | || ||   ::|:  |||||||: || |   |
Db       1811 LAGRSAPERTALLLDLVRTEVAAVLGHGDPAAIGAARTFKDAGFDSLTAVDLRNRLNTRT 1870

Qy         61 GLRLPVSLVFDHPTPAKLA 79
              ||||| :||||||||  ||
Db       1871 GLRLPATLVFDHPTPLALA 1889


RESULT 14
US-08-804-198-3
; Sequence 3, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
```

```
;       REGISTRATION NUMBER:  36,470
;       REFERENCE/DOCKET NUMBER:  P9113
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  317-276-3885
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1996 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-804-198-3

  Query Match             53.8%;  Score 219;  DB 2;  Length 1996;
  Best Local Similarity   58.2%;  Pred. No. 9.4e-21;
  Matches   46;  Conservative   11;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || ::: :: | ||:||| ||||| | || ||   ::|:  ||||||||: || |   |
Db       1811 LAGRSAPERTALLLDLVRTEVAAVLGHGDPAAIGAARTFKDAGFDSLTAVDLRNRLNTRT 1870

Qy         61 GLRLPVSLVFDHPTPAKLA 79
              ||||| :||||||||  ||
Db       1871 GLRLPATLVFDHPTPLALA 1889


RESULT 15
US-08-804-227C-4
; Sequence 4, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3729 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-4

  Query Match             53.8%;  Score 219;  DB 2;  Length 3729;
  Best Local Similarity   57.7%;  Pred. No. 2.3e-20;
  Matches   45;  Conservative   10;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          5 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 64
              |: :  | :|  :| | |||| |    :|  |::|: |||||||||| || |  ||||||
Db       3551 TAQEGHALVLAEIRAHAAAVLGHGSDDSIPEDRAFKDLGFDSLTAVEMRNRLSAATGLRL 3610

Qy         65 PVSLVFDHPTPAKLAVHL 82
              | :|||||||| :|| ||
Db       3611 PATLVFDHPTPGELAGHL 3628


Search completed: June 17, 2003, 13:09:28
Job time : 2.49039 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 4.01696 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_1979_2060
Perfect score: 407
Sequence: 1 LAHQTSADQRAALLELVRDH.....RLPVSLVFDHPTPAKLAVHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	407	100.0	3972	21	AAB23749	S. avermitilis ave
2	407	100.0	3972	22	AAG65264	Streptomyces averm
3	407	100.0	3972	22	AAG65268	Streptomyces averm
4	243	59.7	5532	21	AAB23752	S. avermitilis ave
5	243	59.7	5532	22	AAG65267	Streptomyces averm
6	240	59.0	6797	22	AAB31558	Pimaricin biosynth
7	233	57.2	6239	21	AAB23750	S. avermitilis ave
8	233	57.2	6239	22	AAG65265	Streptomyces averm
9	229	56.3	3739	21	AAB18638	Amino acid sequenc
10	229	56.3	3739	21	AAY77193	S. venezuelae macr
11	229	56.3	3739	21	AAY77201	S. venezuelae pik
12	229	56.3	3739	21	AAY67202	Narbonolide syntha
13	229	56.3	9477	22	AAE10144	Streptomyces nours
14	229	56.3	12199	21	AAY77180	S. venezuelae pik
15	225	55.3	1560	21	AAY67203	Narbonolide syntha
16	225	55.3	1561	21	AAY77202	S. venezuelae pik
17	225	55.3	1562	21	AAB18639	Amino acid sequenc
18	225	55.3	1562	21	AAY77194	S. venezuelae macr
19	224	55.0	1721	19	AAW52847	A. mediterranei ri
20	224	55.0	3816	21	AAY92708	S. antibioticus 8,
21	223	54.8	3724	18	AAW23718	Platenolide syntha
22	223	54.8	3724	18	AAW22608	Platenolide syntha
23	223	54.8	4150	21	AAY92707	S. antibioticus 8,
24	222	54.5	4545	18	AAW22611	Hybrid srmG/tylG O
25	222	54.5	4550	18	AAW23716	Platenolide syntha
26	222	54.5	4550	18	AAW22606	Platenolide syntha
27	221	54.3	5069	19	AAW52846	A. mediterranei ri
28	221	54.3	11096	22	AAE10129	Streptomyces nours
29	219	53.8	1996	18	AAW22607	Platenolide syntha
30	219	53.8	1996	18	AAW23717	Platenolide syntha
31	219	53.8	3729	18	AAW22603	Tylactone synthase
32	217	53.3	4881	21	AAB23751	S. avermitilis ave
33	217	53.3	4881	22	AAG65266	Streptomyces averm
34	216	53.1	1891	19	AAW52844	Amycolatopsis medi
35	216	53.1	3413	19	AAW52849	A. mediterranei ri
36	215	52.8	5435	22	AAE10145	Streptomyces nours
37	212	52.1	1688	19	AAW52848	A. mediterranei ri
38	212	52.1	2152	20	AAY39298	SpnB a polyketide
39	212	52.1	2152	22	AAB70966	S. spinosa protein
40	211	51.8	4630	21	AAY77177	S. venezuelae vep
41	210	51.6	3170	20	AAY39299	SpnC a polyketide
42	210	51.6	3170	22	AAB70967	S. spinosa protein
43	209.5	51.5	186	22	AAB19902	Polyketide synthas
44	209	51.4	7068	22	AAE10142	Streptomyces nours
45	207.5	51.0	3201	22	AAB82214	Polyketide synthas

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 407; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 9.8e-46;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
Db 1979 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 2038

Qy 61 GLRLPVSLVFDHPTPAKLAVHL 82
Db 2039 GLRLPVSLVFDHPTPAKLAVHL 2060

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 407; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 9.8e-46;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
Db 1979 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 2038

Qy 61 GLRLPVSLVFDHPTPAKLAVHL 82
Db 2039 GLRLPVSLVFDHPTPAKLAVHL 2060

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 407; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 9.8e-46;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 2038

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||||||||||||||||||||
Db       2039 GLRLPVSLVFDHPTPAKLAVHL 2060


RESULT 4
AAB23752
ID   AAB23752 standard; Protein; 5532 AA.
XX
AC   AAB23752;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 278-305; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;

  Query Match             59.7%;  Score 243;  DB 21;  Length 5532;
  Best Local Similarity   59.8%;  Pred. No. 4.8e-23;
  Matches   49;  Conservative    5;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || ||    |    || ||| |:| || |   |  | ||::|| |||||||||| || | : |
Db       1756 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 1815

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :| |||| |   |  ||
Db       1816 GLRLPTTLAFDHPNPTTLTHHL 1837


RESULT 5
AAG65267
ID   AAG65267 standard; Protein; 5532 AA.
XX
AC   AAG65267;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 7.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 217-235; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   5532 AA;

  Query Match             59.7%;  Score 243;  DB 22;  Length 5532;
  Best Local Similarity   59.8%;  Pred. No. 4.8e-23;
  Matches   49;  Conservative    5;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || ||    |    || ||| |:| || |   |  | ||::|| |||||||||| || | : |
Db       1756 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 1815

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :| |||| |   |  ||
Db       1816 GLRLPTTLAFDHPNPTTLTHHL 1837


RESULT 6
AAB31558
ID   AAB31558 standard; Protein; 6797 AA.
XX
AC   AAB31558;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase polypeptide.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin.
XX
```

OS Streptomyces natalensis.
XX
PN WO200077222-A1.
XX
PD 21-DEC-2000.
XX
PF 14-JUN-2000; 2000WO-EP06227.
XX
PR 14-JUN-1999; 99EP-0201893.
XX
PA (STAM) DSM NV.
XX
PI Martin JF, Aparicio JF, Colina AJ;
XX
DR WPI; 2001-080693/09.
DR N-PSDB; AAF24892.
XX
PT New polynucleotides encoding enzymes involved in the biosynthesis of
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT the synthesis of new compounds -
XX
PS Disclosure; Page 81-101; 116pp; English.
XX
CC The present sequence represents a polyketide synthase which is associated
CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC is useful for the oxidative modification of a methyl group of a suitable
CC compound, e.g. a bioactive compound including a secondary metabolite,
CC antibiotics and anticancer agents. Recombinant cells comprising the
CC gene are useful for the production of pimaricin. The polyketide synthase
CC polynucleotide may be over expressed in Streptomyces, leading to an
CC increase in the biosynthesis of pimaricin, as a source of primers for
CC amplification reaction and as probes.
XX
SQ Sequence 6797 AA;

Query Match 59.0%; Score 240; DB 22; Length 6797;
Best Local Similarity 59.8%; Pred. No. 1.6e-22;
Matches 49; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

```
Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  ::|::| |||::||  :| || || |: ||||::|: || |||||:| |||| ||| `
Db       4893 LAGLSTAERREALLDVVRAQIATVLGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKAT 4952

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| : |||:||   || ||
Db       4953 GLRLPATTVFDYPTVDALAAHL 4974
```

RESULT 7
AAB23750
ID AAB23750 standard; Protein; 6239 AA.
XX
AC AAB23750;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 32; Page 223-254; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 6239 AA;

Query Match 57.2%; Score 233; DB 21; Length 6239;
Best Local Similarity 61.0%; Pred. No. 1.3e-21;
Matches 50; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

```
Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || |:: ::| ||| ||| |||||| |:    |   ::|| |||||||||| || |  ||
Db       6032 LARQSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAAT 6091

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||  :| || |||| || ||
Db       6092 GLRLRATLAFDFPTPAALAEHL 6113
```

RESULT 8
AAG65265
ID AAG65265 standard; Protein; 6239 AA.
XX
AC AAG65265;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 5.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -

```
XX
PS   Claim 4; Page 180-201; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   6239 AA;

  Query Match             57.2%;  Score 233;  DB 22;  Length 6239;
  Best Local Similarity   61.0%;  Pred. No. 1.3e-21;
  Matches   50;  Conservative    8;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              || |:: ::| ||| ||| |||||| |:    |   ::|| |||||||||| || |  ||
Db       6032 LARQSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAAT 6091

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||  :| || |||| || ||
Db       6092 GLRLRATLAFDFPTPAALAEHL 6113


RESULT 9
AAB18638
ID   AAB18638 standard; Protein; 3739 AA.
XX
AC   AAB18638;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of narbonolide synthase subunit 2 (PICAII).
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 11-12; 117pp; English.
XX
CC   The present sequence represents a narbonolide synthase subunit 2
CC   (PICAII). The nucleotide sequence encoding it is used in the course of
CC   the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence   3739 AA;

  Query Match             56.3%;  Score 229;  DB 21;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 2.4e-21;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  ||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 10
AAY77193
ID   AAY77193 standard; Protein; 3739 AA.
XX
AC   AAY77193;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic enzyme PikAII, SEQ ID NO:33.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87298.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 403-415; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
```

```
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC   Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC   AAZ87295-Z87302.
XX
SQ   Sequence   3739 AA;

  Query Match             56.3%;  Score 229;  DB 21;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 2.4e-21;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 11
AAY77201
ID   AAY77201 standard; Protein; 3739 AA.
XX
AC   AAY77201;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster protein #2.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87318.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAY77200-Y77203 and AAY80997 represent  proteins encoded by the
CC   macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
CC   ATCC 15439, as given in figure 31.
XX
SQ   Sequence   3739 AA;

  Query Match             56.3%;  Score 229;  DB 21;  Length 3739;
  Best Local Similarity   59.8%;  Pred. No. 2.4e-21;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||  |  |: | | :||| ||| || |  |  :  :::||  |||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 12
AAY67202
ID   AAY67202 standard; protein; 3739 AA.
XX
AC   AAY67202;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Narbonolide synthase subunit 2 (PICAII) protein sequence.
XX
KW   Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 2;
KW   PICAII; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS   Streptomyces venezuelae.
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
```

```
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   N-PSDB; AAZ56001.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Example 2; Page 13-14; 98pp; English.
XX
CC   This is the Streptomyces venezuelae narbonolide synthase subunit 2,
CC   PICAII protein sequence. The invention relates to recombinant DNA
CC   containing a coding sequence for a narbonolide polyketide synthase
CC   (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC   a series of condensations and subsequent modifications. Modular PKSs are
CC   responsible for the production of many antibiotics including picromycin.
CC   The narbonolide PKS consists of a loading module, six extender modules,
CC   and two thioester domains. Four proteins make up the narbonolide PKS
CC   (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC   and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC   PICAIII includes extender module 5 and PICAIV includes extender module 6
CC   and a type II thioesterase domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this  conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine.
XX
SQ   Sequence   3739 AA;

  Query Match            56.3%;  Score 229;  DB 21;  Length 3739;
  Best Local Similarity  59.8%;  Pred. No. 2.4e-21;
  Matches   49;  Conservative    8;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
              ||   |  |: | | :||| ||| || |  |  :  :::||  |||||||||| || |  ||
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPVSLVFDHPTPAKLAVHL 82
              ||||| :|||||||| :|| ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 13
AAE10144
ID   AAE10144 standard; Protein; 9477 AA.
XX
AC   AAE10144;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (complete).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          41..464
FT                   /label= KS15_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          578..889
FT                   /label= AT15_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          903..1102
FT                   /label= DH15_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1446..1731
FT                   /label= ER15_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          1740..1988
FT                   /label= KR15_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          2023..2096
FT                   /label= ACP15_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          2117..2538
FT                   /label= KS16_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2635..2953
FT                   /label= AT16_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2967..3167
FT                   /label= DH16_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3257..3500
FT                   /label= KR16_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3539..3612
FT                   /label= ACP16_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3634..4057
FT                   /label= KS17_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4153..4472
FT                   /label= AT17_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4486..4725
FT                   /label= DH17_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4997..5245
FT                   /label= KR17_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5277..5350
FT                   /label= ACP17_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
```

XX
PS Claim 15; Page 255-260; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a PKS type I encoding Streptomyces noursei
CC nystatin gene, NysI (complete).
XX
SQ Sequence 9477 AA;

Query Match 56.3%; Score 229; DB 22; Length 9477;
Best Local Similarity 59.8%; Pred. No. 8.2e-21;
Matches 49; Conservative 5; Mismatches 28; Indels 0; Gaps 0;

Qy 1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
Db 7801 LAEQPAADHDAILLALVTEKAATVLGHADHDLLEPDLPFRDLGFDSLTAVDLRNQLTAET 7860

Qy 61 GLRLPVSLVFDHPTPAKLAVHL 82
Db 7861 GLTLPATLVFDHPNPAALAAHL 7882

RESULT 14
AAY77180
ID AAY77180 standard; Protein; 12199 AA.
XX
AC AAY77180;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87285.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 19; Page 315-353; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents a protein
CC encoded by the macrolide biosynthetic gene cluster (pik) from
CC Streptomyces venezuelae ATCC 15439.
XX
SQ Sequence 12199 AA;

Query Match 56.3%; Score 229; DB 21; Length 12199;
Best Local Similarity 59.8%; Pred. No. 1.2e-20;
Matches 49; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

Qy 1 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 60
Db 8831 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 8890

Qy 61 GLRLPVSLVFDHPTPAKLAVHL 82
Db 8891 GLRLPATLVFDHPTPGELAGHL 8912

RESULT 15
AAY67203
ID AAY67203 standard; protein; 1560 AA.
XX
AC AAY67203;
XX
DT 23-MAR-2000 (first entry)
XX
DE Narbonolide synthase subunit 3 (PICAIII) protein sequence.
XX
KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 3;
KW PICAIII; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS Streptomyces venezuelae.
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR N-PSDB; AAZ56001.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Example 2; Page 14; 98pp; English.
XX
CC This is the Streptomyces venezuelae narbonolide synthase subunit 3,
CC PICAIII protein sequence. The invention relates to recombinant DNA

CC containing a coding sequence for a narbonolide polyketide synthase
CC (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC a series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC transferase enzyme is required for this conversion, and the desosamine
CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC DNA of the invention is used to express, in transformed cells,
CC narbonolide (or its derivatives) or other ketolides (particularly
CC hybrids), which may then be converted (e.g. by other enzymes
CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.
XX
SQ Sequence 1560 AA;

Query Match 55.3%; Score 225; DB 21; Length 1560;
Best Local Similarity 60.0%; Pred. No. 2.6e-21;
Matches 45; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

```
Qy          8 DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVS 67
              ::| ||| ||| | |||| |: |  :|| ::|  |||||||||: || |    | ||| :
Db       1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy         68 LVFDHPTPAKLAVHL 82
                ||||||||  || ||
Db       1460 TVFDHPTPAALAAHL 1474
```

Search completed: June 17, 2003, 13:02:31
Job time : 5.01696 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 8.22676 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_1715_1892
Perfect score: 953
Sequence: 1 GTTLITGGTGALATHLTHHL.....LAHHRHTHHLPATSIAWGTW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	953	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	948	99.5	6239	2	Q9S0R7	Q9s0r7 streptomyce
3	894.5	93.9	3626	2	Q9F7T9	Q9f7t9 streptomyce
4	573	60.1	3970	2	Q93HI8	Q93hi8 streptomyce
5	501	52.6	6315	2	Q9ADL6	Q9adl6 polyangium
6	500	52.5	5435	2	Q9L4X2	Q9l4x2 streptomyce
7	483	50.7	11096	2	Q9L4W3	Q9l4w3 streptomyce
8	481	50.5	5644	2	Q93NX8	Q93nx8 streptomyce
9	480	50.4	1879	2	O30768	O30768 streptomyce
10	479	50.3	1569	2	O30767	O30767 streptomyce
11	475	49.8	2066	2	Q9L4X1	Q9l4x1 streptomyce
12	471.5	49.5	9507	2	Q9EWA1	Q9ewa1 streptomyce
13	471	49.4	2034	2	Q93NX7	Q93nx7 streptomyce
14	471	49.4	3576	16	Q9EX53	Q9ex53 streptomyce
15	471	49.4	4840	2	Q93HJ4	Q93hj4 streptomyce
16	468	49.1	1808	2	Q9EWA2	Q9ewa2 streptomyce
17	467	49.0	2731	2	Q9Z3T9	Q9z3t9 pseudomonas
18	464	48.7	8817	2	Q53840	Q53840 polyangium
19	461.5	48.4	9477	2	Q9L4X3	Q9l4x3 streptomyce
20	457	48.0	1864	2	O33955	O33955 streptomyce
21	457	48.0	10917	2	Q93NW6	Q93nw6 streptomyce
22	456.5	47.9	6797	2	Q9X993	Q9x993 streptomyce
23	454	47.6	3613	2	Q93HJ1	Q93hj1 streptomyce
24	452	47.4	4881	2	Q9S0R3	Q9s0r3 streptomyce
25	452	47.4	6260	2	Q54299	Q54299 streptomyce
26	449	47.1	8563	2	Q54297	Q54297 streptomyce
27	447	46.9	3654	2	O30766	O30766 streptomyce
28	445.5	46.7	9510	2	Q93NX9	Q93nx9 streptomyce
29	445	46.7	2066	2	Q9Z3T8	Q9z3t8 pseudomonas
30	444.5	46.6	6146	2	Q93HJ5	Q93hj5 streptomyce
31	442	46.4	4340	2	O30764	O30764 streptomyce
32	440	46.2	10223	2	Q54296	Q54296 streptomyce
33	439	46.1	4735	2	O54666	O54666 amycolatops
34	436	45.8	3939	2	Q93HJ3	Q93hj3 streptomyce
35	436	45.8	5588	2	Q9ALM2	Q9alm2 saccharopol
36	436	45.8	7525	2	Q9KIE0	Q9kie0 streptomyce
37	435	45.6	4498	2	Q93HE5	Q93he5 streptomyce
38	431	45.2	4151	16	O53490	O53490 mycobacteri
39	429.5	45.1	6420	2	P95814	P95814 streptomyce
40	428.5	45.0	4613	2	Q9ZGI5	Q9zgi5 streptomyce
41	427	44.8	4290	2	Q9WXC0	Q9wxc0 micromonosp
42	425	44.6	3192	2	Q9L4W4	Q9l4w4 streptomyce
43	424.5	44.5	1191	2	Q9RHV0	Q9rhv0 streptomyce
44	423	44.4	4472	2	O33954	O33954 streptomyce
45	422	44.3	1907	2	Q93HI9	Q93hi9 streptomyce

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 953; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 7.1e-77;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1715 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 1774

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1775 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 1834

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1835 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 1892
```

RESULT 2
Q9S0R7
ID Q9S0R7 PRELIMINARY; PRT; 6239 AA.
AC Q9S0R7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 2.
GN AVEA2.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84475.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000169; SHprot_acsite.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 4.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 6239 AA; 666283 MW; A706280940B502FA CRC64;

Query Match 99.5%; Score 948; DB 2; Length 6239;
Best Local Similarity 99.4%; Pred. No. 3.4e-76;
Matches 177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2306 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 2365

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2366 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 2425

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2426 QLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 2483
```

RESULT 3
Q9F7T9
ID Q9F7T9 PRELIMINARY; PRT; 3626 AA.
AC Q9F7T9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Avermectin polyketide synthase (Fragment).
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31271;
RA Hong Y.-S., Lee J.J.;
RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT and polyketide synthase gene from Streptomyces avermitilis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275943; AAG09812.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT NON_TER 3626 3626
SQ SEQUENCE 3626 AA; 380557 MW; 6272F5F088C1A8D0 CRC64;

Query Match 93.9%; Score 894.5; DB 2; Length 3626;
Best Local Similarity 94.4%; Pred. No. 1.2e-71;
Matches 169; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 1740

Qy         61 CDTSNPDQL-QQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
              |||| |       |||||||||:|||||| ||||||||||||||||||||||||||||||
Db       1741 CDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPTQLNNVLRAKAHSAHLL 1800

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 1859
```

RESULT 4
Q93HI8
ID Q93HI8 PRELIMINARY; PRT; 3970 AA.
AC Q93HI8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Modular polyketide synthase.
GN OLMA4.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMBL; AB070940; BAB69199.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000794; Ketoacyl-synt.

```
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3970 AA;  412657 MW;  A93E583FAFA68C47 CRC64;

  Query Match             60.1%;  Score 573;  DB 2;  Length 3970;
  Best Local Similarity   62.9%;  Pred. No. 1.4e-42;
  Matches  112;  Conservative   12;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |    | || |      ||||| ||| | |  |  :|   | |:|:|
Db       1458 GTALITGGTGTLGALLARHLVTRHGVTSLLLTSRRGPHAPGATELVAELTAAGAHVTVTA 1517

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              |||::|:||  || ||| :|||||||||||| ||||| | || |||: ||: |  :|  ||
Db       1518 CDTTDPEQLATLLTTIPTEHPLTTVIHTAGTLDDATTTKLTDTQLHTVLQPKIDAATHLH 1577

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               || | |:| ||||||||    |  |||||||| |||||||| ||  ||||:||| |
Db       1578 HLTLHHPVTTFVLYSSAAGQLGTAGQANYAAANTYLDALAHHRRTHGHPATSLAWGFW 1635


RESULT 5
Q9ADL6
ID   Q9ADL6      PRELIMINARY;      PRT;  6315 AA.
AC   Q9ADL6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Soraphen polyketide synthase A.
GN   SORA.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SO CE26;
RA   Ligon J.M., Hill S., Beck J., Zirkle R., Molnar I., Zawodny J.,
RA   Money S., Schupp T.;
RT   "Characterization of the biosynthetic gene cluster for the antifungal
RT   polyketide soraphen A from Sorangium cellulosum So ce26.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U24241; AAK19883.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_4.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6315 AA;  671301 MW;  7B3FB977006C12FE CRC64;

  Query Match             52.6%;  Score 501;  DB 2;  Length 6315;
  Best Local Similarity   53.4%;  Pred. No. 7.8e-36;
  Matches   95;  Conservative   25;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |    :  ||       :|||||| | | | |:   |:|:  |  :|:
Db       3803 GTVLITGGTGTLGALVARHLVARHGAKHLLLTSRQGAHAPGAEASRTELEALGASVTLRA 3862

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  || ||::||  |||| |:| || |||  |  ::| ::: |    |  :|  ||
Db       3863 CDAADPRALQALLDSIPSAHPLTAVVHAAGALDDGLLGAMSPERIDRVFAPKLDAAWHLH 3922

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||  || ||||:||||   |:|||:|||||||:||||||| | |||:|:||| |
Db       3923 ELTQDKPLAAFVLFSSAAGVLGSPGQSNYAAANAFLDALAHHRRAHGLPASSLAWGYW 3980


RESULT 6
Q9L4X2
ID   Q9L4X2      PRELIMINARY;      PRT;  5435 AA.
AC   Q9L4X2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysJ.
GN   NYSJ.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71767.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5435 AA;  562659 MW;  AA55465DF087A38C CRC64;

  Query Match             52.5%;  Score 500;  DB 2;  Length 5435;
  Best Local Similarity   55.1%;  Pred. No. 8e-36;
  Matches   98;  Conservative   16;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| ||  |  || |     :|||| | ||  | | |  | :|   |  :|:
Db       1740 GTVLLTGGTGGLAGLLARHLVTEHGARHLLLAGRRGPDAPGAAALHAELTALGAEVTVAA 1799
```

```
Qy          61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || ::    |  || |:| :|||| |:||||:||| ||| | | :|  ||| |  :|  ||
Db        1800 CDVADRTALAALLATVPAEHPLTAVVHTAGVLDDGTLTALNPDRLATVLRPKVDAAWHLH 1859

Qy         121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               ||:|  | ||||||| |   | ||||||| | :||||| |||   |||||:||| |
Db        1860 DLTRHLDLAAFVLYSSTAGVMGGPGQANYAAGNTFLDALAAHRHALGLPATSLAWGAW 1917


RESULT 7
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match           50.7%;  Score 483;  DB 2;  Length 11096;
  Best Local Similarity  54.5%;  Pred. No. 6.4e-34;
  Matches   97;  Conservative   19;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |   |  ||      :|||| || ||    | || :|   |  : |
Db       1381 GTVLITGGTGGLGAVLARHLVATHGVRHLLLASRRGPAADGADDLTAELTGLGATVHIAA 1440

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  |  || |:|  |||| |:||||::||  | :||| :|: ||| || :|  ||
Db       1441 CDVADPAALADLLGTVPAGHPLTVVVHTAGVVDDGVLGSLTPQRLDTVLRPKADAAWHLH 1500

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |:|  | ||||:|| ||| |:||||||| ||:|||||  |    |||||:||| |
Db       1501 EATRHLDLDAFVLFSSVAATLGSPGQANYAAGNAFLDALAARRAATGLPATSLAWGPW 1558


RESULT 8
Q93NX8
ID   Q93NX8      PRELIMINARY;      PRT;  5644 AA.
AC   Q93NX8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphJ.
GN   AMPHJ.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73502.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW   Phosphopantetheine.
SQ   SEQUENCE   5644 AA;  585775 MW;  CF6A35E3BA0D2658 CRC64;

  Query Match           50.5%;  Score 481;  DB 2;  Length 5644;
  Best Local Similarity  51.7%;  Pred. No. 4.3e-34;
  Matches   92;  Conservative   24;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| ||: |  ||      :|||||| ||  | |  |  | : |    :
Db       1420 GTVLVTGGTGGLASALARHLVAEHDIRHLLLTSRRGPDAPGAADLVQALAELGAEARVAA 1479

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : |  || ::| :|||| |:||||:|||    :||| :|::|:| |  :|  ||
Db       1480 CDVADREALAALLASVPAEHPLTAVVHTAGVLDDGIFPSLTPDRLDSVMRPKVDAAWHLH 1539

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||:  | ||||||| |   |:||||||| | :||||| || :  |||||:||| |
Db       1540 DLTRDLDLAAFVLYSSTAGVMGSPGQANYAAGNTFLDALAAHRQSLGLPATSLAWGAW 1597


RESULT 9
O30768
ID   O30768      PRELIMINARY;      PRT;  1879 AA.
AC   O30768;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase module 7.
GN   NIDA5.
OS   Streptomyces caelestis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=36816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL-2821;
RX   MEDLINE=98053867; PubMed=9393718;
RA   Kakavas S.J., Katz L., Stassi D.;
RT   "Identification and characterization of the niddamycin polyketide
RT   synthase genes from Streptomyces caelestis.";
RL   J. Bacteriol. 179:7515-7522(1997).
DR   EMBL; AF016585; AAC46028.1; -.
```

```
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   1879 AA;  199133 MW;  6B84CFDC5F425BA7 CRC64;

  Query Match             50.4%;  Score 480;  DB 2;  Length 1879;
  Best Local Similarity   54.7%;  Pred. No. 1.5e-34;
  Matches   98;  Conservative   19;  Mismatches   60;  Indels    2;  Gaps    2;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||||| ||:   |  |:  :||:|||| ||: | |  |   | | | | |||
Db       1212 GTILITGGTGALGTHIATWL-AHKGAKHLILTSRQGPNAPGATQLQHHLTQLGTHTTITP 1270

Qy         61 CDTSNPDQLQQLLNTIPPQH-PLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
              || ::| |:: |:| |   | ||  : ||||      ||:|||||  ::| ||   | ||
Db       1271 CDLTDPQQIENLINHINTHHPPLQAIFHTAGTPQTTPLTHLTPTQATDILNAKTTGAELL 1330

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |  | ||||   | :||||||:|: ||: ||||||:||||| | ||   | ||||||
Db       1331 HHHTTHTPLHTVVHFSSAAATWGSSGQSLYAAANAHLDALAQHTHTQGQPTTSIAWGVW 1389


RESULT 10
O30767
ID   O30767      PRELIMINARY;      PRT;  1569 AA.
AC   O30767;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase module 6.
GN   NIDA4.
OS   Streptomyces caelestis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=36816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL-2821;
RX   MEDLINE=98053867; PubMed=9393718;
RA   Kakavas S.J., Katz L., Stassi D.;
RT   "Identification and characterization of the niddamycin polyketide
RT   synthase genes from Streptomyces caelestis.";
RL   J. Bacteriol. 179:7515-7522(1997).
DR   EMBL; AF016585; AAC46027.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   1569 AA;  166676 MW;  91E6670FB71CC670 CRC64;

  Query Match             50.3%;  Score 479;  DB 2;  Length 1569;
  Best Local Similarity   54.7%;  Pred. No. 1.5e-34;
  Matches   98;  Conservative   19;  Mismatches   60;  Indels    2;  Gaps    2;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||||| ||:   |  |:  :||:|||| ||: | |  |   | | | | |||
Db       1165 GTILITGGTGALGTHIATWL-AHKGAKHLILTSRQGPNAPGATQLQHHLTQLGTHTTITP 1223

Qy         61 CDTSNPDQLQQLLNTIPPQH-PLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
              || ::| |:: |:| |   | ||  : ||||      ||:|||||  ::| ||   | ||
Db       1224 CDLTDPQQIENLINHINTHHPPLQAIFHTAGTPQTTPLTHLTPTQATDILNAKTTGAELL 1283

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |  | ||||   | :||||||:|: ||: ||||||:||||| | ||   | ||||||
Db       1284 HHHTTHTPLHTVVHFSSAAATWGSSGQSLYAAANAHLDALAQHTHTQGQPTTSIAWGLW 1342


RESULT 11
Q9L4X1
ID   Q9L4X1      PRELIMINARY;      PRT;  2066 AA.
AC   Q9L4X1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NysK.
GN   NYSK.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71768.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2066 AA;  214572 MW;  CB02F5991FB2B56E CRC64;

  Query Match             49.8%;  Score 475;  DB 2;  Length 2066;
  Best Local Similarity   52.8%;  Pred. No. 4.6e-34;
  Matches   94;  Conservative   17;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |   |  || | :  :|||| | ||  | |  |  :|  :|  : :
Db       1416 GTVLITGGTGGLGAALARHLVTVRGARHLLLAGRRGPEAPGAGELVAELTAQGADVRVAA 1475

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||  :   |  || |:|  |||| |:||||:|||| : :||| ||  ||| || :|  ||
Db       1476 CDVGDRTALDALLATVPAAHPLTAVVHTAGVLDDALIGSLTPDQLATVLRPKADAAWHLH 1535

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               |:   |  |||||| :   |:||| |||||||||||||| ||    ||| |:||| |
Db       1536 DATRGLDLAGFVLYSSVSGVLGSPGQGNYAAANAYLDALARHRADQGLPALSLAWGPW 1593


RESULT 12
```

```
Q9EWA1
ID   Q9EWA1      PRELIMINARY;      PRT;  9507 AA.
AC   Q9EWA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PimS2 protein.
GN   PIMS2.
OS   Streptomyces natalensis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=68242;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20547809; PubMed=11094342;
RA   Aparicio J.F., Fouces R., Mendes M.V., Olivera N., Martin J.F.;
RT   "A complex multienzyme system encoded by five polyketide synthase
RT   genes is involved in the biosynthesis of the 26-membered polyene
RT   macrolide pimaricin in Streptomyces natalensis.";
RL   Chem. Biol. 7:895-905(2000).
DR   EMBL; AJ278573; CAC20921.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine.
SQ   SEQUENCE   9507 AA;  994228 MW;  57BD80C8AD37E89F CRC64;

  Query Match             49.5%;  Score 471.5;  DB 2;  Length 9507;
  Best Local Similarity   51.7%;  Pred. No. 5.8e-33;
  Matches   92;  Conservative   21;  Mismatches   64;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||||  |:    |      :|||| || |   | |  |  :: : |  :|:
Db       6025 GTVLVTGGTGALGGHVARWL-AEAGAEHLLLVSRRGTDAPGAGELAAEIAELGARVTVAA 6083

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              |||:: | |   :|  || :|||| | |||| :|| ||  ||| |  :||| |  :  ||
Db       6084 CDTADRDALAAVLAAIPDEHPLTAVFHTAGTVDDGTLDTLTPEQFASVLRTKVTATRNLH 6143

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |:   |:||||:|| | | ||||| |||| ||:||| | ||  | |||||:||| |
Db       6144 ETTREQDLSAFVLFSSVAGTLGAPGQGNYAAGNAFLDAFAAHRRAHGLPATSLAWGPW 6201


RESULT 13
Q93NX7
ID   Q93NX7      PRELIMINARY;      PRT;  2034 AA.
AC   Q93NX7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphK.
GN   AMPHK.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73503.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   2034 AA;  212063 MW;  83E962B7DC3D5747 CRC64;

  Query Match             49.4%;  Score 471;  DB 2;  Length 2034;
  Best Local Similarity   52.2%;  Pred. No. 1e-33;
  Matches   93;  Conservative   19;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |   |  || | :  :|||| || ||  |   |  :|   |  :|::
Db       1386 GTVLITGGTGGLGAALARHLVTSRGARHLLLASRRGPDAPGTSELVAELTGLGAQVTVSA 1445

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||  : | :  |: ::| :|||| |:||||:||||   :||| ||  ||| || :|  ||
Db       1446 CDVGDRDAVDTLVASVPAEHPLTAVVHTAGVLDDALTGSLTPEQLAGVLRPKADAALHLH 1505

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |    | ||||||| :  | ||||||||||| |||||| |    ||  |:||| |
Db       1506 EATLGQDLAAFVLYSSISGVIGGPGQANYAAANASLDALAHRRRAAGLPGLSLAWGPW 1563


RESULT 14
Q9EX53
ID   Q9EX53      PRELIMINARY;      PRT;  3576 AA.
AC   Q9EX53;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative type I polyketide synthase.
GN   SCO6274 OR SC2C4.04C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
```

```
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL512902; CAC22145.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR003662; sub_transporter.
DR   InterPro; IPR001998; Xylose_isom.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   TIGRFAMs; TIGR00128; fabD; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR   PROSITE; PS00173; XYLOSE_ISOMERASE_2; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3576 AA;  377350 MW;  8A83C6F90AC9928B CRC64;

  Query Match             49.4%;  Score 471;  DB 16;  Length 3576;
  Best Local Similarity   53.9%;  Pred. No. 2e-33;
  Matches   96;  Conservative   20;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||| |||| | |  || |   |:|||| || | : | || |  :|    |  :|||
Db       1418 GTVLITGATGALGTVLARHLVTQHGTRHLLLLSRRGTNAPGAQELQHELTALGATVTITA 1477

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::   |  :|: || :|||| |:||||:|||  |::||  :| :||| |  :|  ||
Db       1478 CDAADRHALTSVLDDIPAEHPLTAVVHTAGVLDDGLLSSLTEERLADVLRPKVDAAWNLH 1537

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||||| |   |  |||||||| | :||||| ||    ||| |:||| |
Db       1538 ELTKDHDLEAPVLYSSVAGLIGNAGQANYAAGNTFLDALAQHRRALGLPAASLAWGLW 1595


RESULT 15
Q93HJ4
ID   Q93HJ4      PRELIMINARY;      PRT;  4840 AA.
AC   Q93HJ4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   OlmA2 protein.
GN   OLMA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69193.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   4840 AA;  504325 MW;  169BDA3589AD284F CRC64;

  Query Match             49.4%;  Score 471;  DB 2;  Length 4840;
  Best Local Similarity   50.6%;  Pred. No. 2.9e-33;
  Matches   90;  Conservative   23;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ::|| || : | :  || |    :|||| || || | | |   | : |   | |
Db       4401 GTVVVTGATGGIGTEIVRHLATAHGVRHLLLLSRRGPDDPRAAELGRVLAEAGADATFTA 4460

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              |||::   |   |  :| :||:| |:| ||::||  || ||| ::: ||| || :|  ||
Db       4461 CDTADRAALADALADVPAEHPVTAVVHIAGVVDDGVLTTLTPDRIDTVLRPKAEAAQHLH 4520

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||    |: |||:||   |||  |||||||||:|||||| |    |||||:||| |
Db       4521 ELTADLGLSHFVLFSSGVGTFGGAGQANYAAANAFLDALAHRRRAADLPATSLAWGLW 4578


Search completed: June 17, 2003, 13:07:54
Job time : 9.22676 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 1.94115 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_1715_1892
Perfect score: 953
Sequence: 1 GTTLITGGTGALATHLTHHL.....LAHHRHTHHLPATSIAWGTW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	391	41.0	3567	1	ERY2_SACER	Q03132 saccharopol
2	389.5	40.9	3491	1	ERY1_SACER	Q03131 saccharopol
3	326	34.2	3172	1	ERY3_SACER	Q03133 saccharopol
4	295	31.0	3519	1	OL56_STRAT	Q07017 streptomyce
5	271.5	28.5	2110	1	MCAS_MYCBO	Q02251 mycobacteri
6	217	22.8	1774	1	MSAS_PENPA	P22367 penicillium
7	217	22.8	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	203	21.3	4447	1	PKSK_BACSU	P40803 bacillus su
9	201.5	21.1	2505	1	FAS_RAT	P12785 rattus norv
10	196.5	20.6	1538	1	PPSB_MYCTU	Q10978 mycobacteri
11	195	20.5	4427	1	PKSL_BACSU	Q05470 bacillus su
12	191.5	20.1	2511	1	FAS_CHICK	P12276 gallus gall
13	182.5	19.2	2504	1	FAS_HUMAN	P49327 homo sapien
14	179.5	18.8	838	1	FAS_MOUSE	P19096 mus musculu
15	146	15.3	4273	1	PKSM_BACSU	P40872 bacillus su
16	111	11.6	247	1	YGFF_ECOLI	P52037 escherichia
17	109.5	11.5	255	1	FABG_MYCAV	O07399 mycobacteri
18	108.5	11.4	247	1	FABG_MYCTU	Q48930 mycobacteri
19	106	11.1	247	1	YD50_MYCTU	Q11020 mycobacteri
20	105.5	11.1	260	1	TRN2_DATST	P50163 datura stra
21	101.5	10.7	247	1	FABG_CHLTR	P38004 chlamydia t
22	100.5	10.5	268	1	TRNH_DATST	P50165 datura stra
23	99.5	10.4	244	1	FABG_SALTY	O85141 salmonella
24	98.5	10.3	244	1	FABG_ECOLI	P25716 escherichia
25	97	10.2	246	1	PHBB_CHRVI	P45375 chromatium
26	96.5	10.1	246	1	FABG_BACSU	P51831 bacillus su
27	96	10.1	606	1	HM1D_DROAN	P22544 drosophila
28	94.5	9.9	255	1	FABG_MYCSM	P71534 mycobacteri
29	93.5	9.8	260	1	TRN2_HYONI	P50164 hyoscyamus
30	92.5	9.7	244	1	FABG_VIBCH	Q9kqh7 vibrio chol
31	91.5	9.6	241	1	FABG_RICPR	P50941 rickettsia
32	91.5	9.6	245	1	NODG_RHIS3	P72332 rhizobium s
33	90.5	9.5	334	1	GALE_BACHD	Q9kdv3 bacillus ha
34	90	9.4	372	1	3BH3_RAT	P27364 rattus norv
35	89.5	9.4	245	1	NODG_RHIME	P06234 rhizobium m
36	89.5	9.4	246	1	NODG_AZOBR	P17611 azospirillu
37	89.5	9.4	252	1	YCIK_ECOLI	P31808 escherichia
38	89	9.3	244	1	FABG_BUCAI	P57432 buchnera ap
39	89	9.3	609	1	OPA_DROME	P39768 drosophila
40	89	9.3	1301	1	PTP9_DROME	P35832 drosophila
41	88	9.2	247	1	FABG_PSEAE	O54438 pseudomonas
42	88	9.2	906	1	FOX2_CANTR	P22414 candida tro
43	87.5	9.2	242	1	FABG_HAEIN	P43713 haemophilus
44	87.5	9.2	248	1	FABG_CHLMU	Q9pkf7 chlamydia m
45	87.5	9.2	1829	1	Y296_HUMAN	O15015 homo sapien

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             41.0%;  Score 391;  DB 1;  Length 3567;
  Best Local Similarity   46.1%;  Pred. No. 5.6e-26;
  Matches   82;  Conservative   24;  Mismatches   70;  Indels    2;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |   |  || |    :|||| || |   | :  |  :::  |   |
Db       3141 GTVLITGGTGTLGRLLARHLVTEHGVRHLLLVSRRGADAPGSDELRAEIEDLGASAEIAA 3200

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              |||:: | |  ||: :|   ||| |:| ||:| |  :|::    :  ||||| :|  ||
Db       3201 CDTADRDALSALLDGLP--RPLTGVVHAAGVLADGLVTSIDEPAVEQVLRAKVDAAWNLH 3258

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|| :| |: |||:||||:    |||  |||||  |:|||  | |  ||| :: || |
Db       3259 ELTANTGLSFFVLFSSAASVLAGPGQGVYAAANESLNALAALRRTRGLPAKALGWGLW 3316


RESULT 2
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
```

```
FT   DOMAIN      1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN      2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN      3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN      3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE     145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING      447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE     677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     1614   1660       NADP.
FT   BINDING     1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     3058   3104       NADP.
FT   BINDING     3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             40.9%;  Score 389.5;  DB 1;  Length 3491;
  Best Local Similarity   44.4%;  Pred. No. 7.4e-26;
  Matches   79;  Conservative   25;  Mismatches   73;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| :   :    |   :   |||| ||:||    |  |  :|:  |   |:
Db       1613 GTVLVTGGTGGVGGQIARWL-ARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAA 1671

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : :::||  |    ||: | | |  ||| |:  ||  ::   |||  |  ||
Db       1672 CDVTDRESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLH 1731

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   ||||||:|| |: |||||   ||  ||||| ||  | :  ||||::|||||
Db       1732 ELTRELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTW 1789


RESULT 3
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN         1   1484       MODULE 5.
FT   DOMAIN      1485   3172       MODULE 6.
FT   DOMAIN        37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN       554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN      1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN      1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN      2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN      2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN      2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN      2926   3172       THIOESTERASE.
FT   ACT_SITE     199    199       THIOESTER BOND.
FT   ACT_SITE     643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     1118   1164       NADP.
FT   BINDING     1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    1661   1661       THIOESTER BOND.
FT   ACT_SITE    2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     2557   2605       NADP.
FT   BINDING     2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT     231    231       S -> SA (IN REF. 2).
FT   CONFLICT     240    240       MISSING (IN REF. 2).
FT   CONFLICT     289    289       A -> R (IN REF. 2).
FT   CONFLICT     493    493       P -> R (IN REF. 2).
FT   CONFLICT     493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                 RWPPAAAPWEQ (IN REF. 1).
```

```
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             34.2%;  Score 326;  DB 1;  Length 3172;
  Best Local Similarity   42.5%;  Pred. No. 2.4e-20;
  Matches   76;  Conservative   23;  Mismatches   74;  Indels    6;  Gaps    4;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||||| |:  ||      : |:| || |   | |  |  :|   |  |||
Db       2557 GTALVTGGTGALGGHVARHL-ARCGVEDLVLVSRRGVDAPAAAELEAELVALGPKTTITA 2615

Qy         61 CDTSNPDQLQQLLNTIPPQ-HPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
              || :: :|| :||  :  |  |: ||:||||: :   |  :   :| :|  ||   | ||
Db       2616 CDVADREQLSKLLEELRGQGRPVRTVVHTAGVPESRPLHEI--GELESVCAAKVTGARLL 2673

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               :|       |||:|| |  :|:     |:||||||||||| |      |||:||| |
Db       2674 DELCPDA--ETFVLFSSGAGVWGSANLGAYSAANAYLDALAHRRRAEGRAATSVAWGAW 2730


RESULT 4
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             31.0%;  Score 295;  DB 1;  Length 3519;
  Best Local Similarity   39.7%;  Pred. No. 1.4e-17;
  Matches   71;  Conservative   26;  Mismatches   78;  Indels    4;  Gaps    3;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQ-QKGIHLTIT 59
              || |:||| | |  |     |        |::| || |   | |  |  :|:   |  :::
Db       1202 GTVLVTGGLGGLGAHTARWL-VGGGADHVVLVSRRGGSAPGAGDLVRELEGLGGARVSVR 1260

Qy         60 TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
              || ::    |: ||: :    |:| | | ||:     |  ::  :  :|: ||   | |
Db       1261 ACDVADRVALRALLSDL--GEPVTAVFHAAGVPQSTPLAEISVQEAADVMAAKVAGAVNL 1318

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               :|     | ||||:|| |  :|: ||| ||||||:|||||  |    |||||:||| |
Db       1319 GELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPATSVAWGMW 1377


RESULT 5
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
```

```
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             28.5%;  Score 271.5;  DB 1;  Length 2110;
  Best Local Similarity   34.6%;  Pred. No. 8.7e-16;
  Matches   62;  Conservative   33;  Mismatches   81;  Indels    3;  Gaps    3;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              |: :|||| | |       |      ::||:|: |: | |:     |:  |  : :
Db       1764 GSYIITGGLGGLGLFFASKLAA-AGCGRIVLTARSQPN-PKARQTIEGLRAAGADIVVEC 1821

Qy         61 CDTSNPDQLQQLLNTIPPQH-PLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
                : : ||   :|::       ||  |:|:| :::||||||:|   ::     |   :  |
Db       1822 GNIAEPDTADRLVSAATATGLPLRGVLHSAAVVEDATLTNITDELIDRDWSPKVFGSWNL 1881

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |: |   ||  | |:|| ||  |:||  |||||:::|  ||      || ::|||| |
Db       1882 HRATLGQPLDWFCLFSSGAALLGSPGQGAYAAANSWVDVFAHWPRAQGLPVSAIAWGAW 1940


RESULT 6
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;
```

```
  Query Match             22.8%;  Score 217;  DB 1;  Length 1774;
  Best Local Similarity   30.1%;  Pred. No. 4.2e-11;
  Matches   58;  Conservative   28;  Mismatches   89;  Indels   18;  Gaps    3;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQ---------- 50
              || ||||| | |    :   |      : ||:: |  |       ::  ||
Db       1412 GTYLITGGLGVLGLEVADFLVEKGARRLLLISRRALPPRRTWDQVSEDLQPTIAKIRLLE 1471

Qy         51 QKGIHLTITTCDTSNPDQLQQLLN-----TIPPQHPLTTVIHTAGILDDATLTNLTPTQL 105
               :|  : :   | : || ::||       ::|    :  |:| ||:||:  :   |
Db       1472 SRGASVHVLPLDITKPDAVEQLTTALDRLSLP---SVQGVVHAAGVLDNELVMQTTRDAF 1528

Qy        106 NNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHT 165
              | ||  |   |  ||::    :  ||::||     |  |||:| : ||:|| || ||
Db       1529 NRVLAPKIAGALALHEVFPPKSVDFFVMFSSCGNLVGFTGQASYGSGNAFLDTLATHRAR 1588

Qy        166 HHLPATSIAWGTW 178
                 | | | :|
Db       1589 LGDAAVSFQWTSW 1601
```

```
RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;
```

```
  Query Match             22.8%;  Score 217;  DB 1;  Length 1876;
  Best Local Similarity   32.5%;  Pred. No. 4.5e-11;
  Matches   62;  Conservative   25;  Mismatches   86;  Indels   18;  Gaps    5;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTP----HAQHLTTQ----------L 49
              ||||| |||   :   |  :    |:|| || |  |         | |:         |
Db       1494 LITGGLGALGLLMADWL-ADRGAHRLVLTGRT-PLPPRRDWQLDTLDTELRRRIDAIRAL 1551

Qy         50 QQKGIHLTITTCDTSNPDQLQQLL--NTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNN 107
              : :|: :     |    : :| ||         |:  :|| ||| :|  :|::|   :
Db       1552 EMRGVTVEAVAADVGCREDVQALLAARDRDGAAPIRGIIHAAGITNDQLVTSMTGDAVRQ 1611

Qy        108 VLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHH 167
              |:  |   : :||       :  | | :|||  || ||| :|||||:||||||  |
Db       1612 VMWPKIGGSQVLHDAFPPGSVDFFYLTASAAGIFGIPGQGSYAAANSYLDALARARRQQG 1671

Qy        168 LPATSIAWGTW 178
                 |: |  |
Db       1672 CHTMSLDWVAW 1682
```

```
RESULT 8
PKSK_BACSU
ID   PKSK_BACSU     STANDARD;      PRT;  4447 AA.
AC   P40803;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksK (PKS).
GN   PKSK.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=95219083; PubMed=7704258;
RA   Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;
RT   "Sequence around the 159 degree region of the Bacillus subtilis
RT   genome: the pksX locus spans 33.6 kb.";
RL   Microbiology 141:299-309(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
```

```
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11039; AAA85144.1; -.
DR   EMBL; Z99112; CAB13590.1; -.
DR   EMBL; Z99113; CAB13601.1; -.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10930; pksK.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF00668; Condensation; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;
KW   Complete proteome.
FT   DOMAIN       1     68       ACYL CARRIER (ACP) 1.
FT   DOMAIN    1063   1130       ACYL CARRIER (ACP) 2.
FT   DOMAIN    2516   2589       ACYL CARRIER (ACP) 3.
FT   DOMAIN    2618   2687       ACYL CARRIER (ACP) 4.
FT   DOMAIN    3868   3937       ACYL CARRIER (ACP) 5.
FT   BINDING     31     31       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   1093   1093       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   2552   2552       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING   2650   2650       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE  2915   2915       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING   3900   3900       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE  4147   4147       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
SQ   SEQUENCE   4447 AA;  496058 MW;  9C6DB4A6C0A9C057 CRC64;

  Query Match             21.3%;  Score 203;  DB 1;  Length 4447;
  Best Local Similarity   30.7%;  Pred. No. 2.2e-09;
  Matches   61;  Conservative   28;  Mismatches   80;  Indels   30;  Gaps    7;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSR--TGPHTPHAQHLTT-----------QL 49
              ||||||  :      |       : |:|| |     |    |:  |:           :|
Db       3579 LITGGTRGIGLLCARHFAECYGVKKLVLTGREQLPPREEWARFKTSNTSLAEKIQAVREL 3638

Qy         50 QQKGIHLTITTCDTSNPDQLQQLLNTIPPQH------PLTTVIHTAGILDDATLTNL--T 101
              : ||: : : :   |:  |::| |      ||      |:  ||| ||: |  ||  :  |
Db       3639 EAKGVQVEMLSLTLSDDAQVEQTL-----QHIKRTLGPIGGVIHCAGLTDMDTLAFIRKT 3693

Qy        102 PTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFG--APGQANYAAANAYLDAL 159
                :  ||  |      |::   : ||  |||:|| :|     : |||:|| ||:|:|
Db       3694 SDDIQRVLEPKVSGLTTLYRHVCNEPLQFFVLFSSVSAIIPELSAGQADYAMANSYMDYF 3753

Qy        160 AHHRHTHHLPATSIAWGTW 178
              |   | | |  |: |  |
Db       3754 A-EAHQKHAPIISVQWPNW 3771


RESULT 9
FAS_RAT
ID   FAS_RAT         STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
```

```
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match             21.1%;  Score 201.5;  DB 1;  Length 2505;
  Best Local Similarity   29.7%;  Pred. No. 1.5e-09;
  Matches   52;  Conservative   34;  Mismatches   86;  Indels    3;  Gaps    2;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 63
              :|||| |      |   |   :  | |:|||:|  | :      : :::|||: ::| :
Db       1882 IITGGLGGFGLELARWLVL-RGAQRLVLTSRSGIRTGYQAKHVREWRRQGIHVLVSTSNV 1940

Qy         64 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLT 123
              |: :  : |:       |:  | : | :| || | | ||    :| : | :    | : |
Db       1941 SSLEGARALIAEATKLGPVGGVFNLAMVLRDAMLENQTPELFQDVNKPKYNGTLNLDRAT 2000

Qy        124 QHT--PLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 176
              :    |  || :|| :   |  ||:||  ||: :: :   |    ||  :: ||
Db       2001 REACPELDYFVAFSSVSCGRGNAGQSNYGFANSTMERICEQRRHDGLPGLAVQWG 2055


RESULT 10
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
```

```
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             20.6%;  Score 196.5;  DB 1;  Length 1538;
  Best Local Similarity   29.0%;  Pred. No. 2.2e-09;
  Matches   51;  Conservative   28;  Mismatches   92;  Indels    5;  Gaps    2;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 63
              |: | || :  ||   |        : :  : |       ||  |   | |  .   |
Db       1155 LVIGATGNIGPHLIRQLARMGAKTIVAMARKPGA----LDELTQCLAATGTDLIAVAADA 1210

Qy         64 SNPDQLQQLLNTIPPQHPLTTVIHTAGILD-DATLTNLTPTQLNNVLRAKAHSAHLLHQL 122
              ::|  :| | :    : |    |: |       | |: :|   :  : | |  : .|||:
Db       1211 TDPAAMQTLFDRFGTELPPLEGIYLAAFAGRPALLSEMTDDDVTTMFRPKLDALALLHRR 1270

Qy        123 TQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
                :  :|:  |||:|| :   |:   |:| | :|:||: |  | |  |||| : || |
Db       1271 SLKSPVRHFVLFSSVSGLLGSRWLAHYTATSAFLDSFAGARRTMGLPATVVDWGLW 1326


RESULT 11
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
```

```
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;

  Query Match             20.5%;  Score 195;  DB 1;  Length 4427;
  Best Local Similarity   28.6%;  Pred. No. 1.1e-08;
  Matches   53;  Conservative   29;  Mismatches   95;  Indels    8;  Gaps    3;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              |  ||||| | |        :         ::|| |: |    :     ||: ||
Db       1410 GVYLITGGAGGLGFIFATEIANQTNDAVVILTGRS-PLDERKKKKLKALQKLGIQAIYRQ 1468

Qy         61 CDTSNPDQLQQLL-NTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
               | ::   :  ||  |      |  :||:||::.|  :      ::  ||  |      |
Db       1469 ADLADKQTVDALLKETQNVYGDLDGIIHSAGLIKDNFIMKKKKEEVQTVLAPKVAGLIHL 1528

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHR------HTHHLPATSI 173
             : |:  ||  |:|:|| |   |: |||:|| |||:::| : :|      |  :    |:
Db       1529 DEATKDIPLDFFILFSSGAGAVGSAGQADYAMANAFMNAFSEYRNGQAELHKRYGKTLSV 1588

Qy        174 AWGTW 178
               | |
Db       1589 CWPLW 1593


RESULT 12
FAS_CHICK
ID   FAS_CHICK      STANDARD;  •   PRT;  2511 AA.
AC   P12276;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=White leghorn; TISSUE=Liver;
RX   MEDLINE=95031085; PubMed=7944406;
RA   Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Amino-terminal blocking group and sequence of the animal fatty acid
RT   synthase.";
RL   Arch. Biochem. Biophys. 314:45-49(1994).
RN   [2]
RP   SEQUENCE OF 75-1775 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89282777; PubMed=2734291;
RA   Holzer K.P., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of chicken liver fatty acid
RT   synthase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN   [3]
RP   SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89139426; PubMed=2917973;
RA   Chirala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,
RA   Wakil S.J.;
RT   "A novel cDNA extension procedure. Isolation of chicken fatty acid
RT   synthase cDNA clones.";
RL   J. Biol. Chem. 264:3750-3757(1989).
RN   [4]
RP   SEQUENCE OF 1752-2512 FROM N.A.
RX   MEDLINE=88320436; PubMed=2842766;
RA   Yuan Z., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of DNA complementary to chicken
RT   liver fatty acid synthase mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN   [5]
RP   SEQUENCE OF 2202-2512 FROM N.A.
RX   MEDLINE=89088152; PubMed=3207710;
RA   Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT   "Characterization of a genomic and cDNA clone coding for the
RT   thioesterase domain and 3' noncoding region of the chicken liver
RT   fatty acid synthase gene.";
RL   Biochemistry 27:7778-7785(1988).
RN   [6]
RP   SEQUENCE OF 2121-2209.
RX   MEDLINE=89192401; PubMed=2648999;
RA   Huang W.-Y., Stoops J.K., Wakil S.J.;
RT   "Complete amino acid sequence of chicken liver acyl carrier protein
RT   derived from the fatty acid synthase.";
RL   Arch. Biochem. Biophys. 270:92-98(1989).
RN   [7]
RP   SEQUENCE OF 2209-2508.
RC   STRAIN=White leghorn;
RX   MEDLINE=89088151; PubMed=3207709;
RA   Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Complete amino acid sequence of the thioesterase domain of chicken
RT   liver fatty acid synthase.";
RL   Biochemistry 27:7773-7777(1988).
RN   [8]
RP   SEQUENCE OF 667-674 AND 1698-1709.
RX   MEDLINE=89323081; PubMed=2751995;
RA   Chang S.I., Hammes G.G.;
RT   "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
RT   fluorescence resonance energy transfer in chicken liver fatty acid
RT   synthase.";
RL   Biochemistry 28:3781-3788(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
CC       POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J04485; AAB46389.1; -.
DR   EMBL; J03860; AAA48767.1; -.
DR   EMBL; J02839; AAA82106.1; ALT_SEQ.
DR   PIR; A33918; XYCHFA.
DR   PIR; A32015; A32015.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;
KW   Alternative splicing; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   DOMAIN        1   ?411       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      427    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1638   1866       ENOYL REDUCTASE.
FT   DOMAIN     1867   2119       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2124   2180       ACYL CARRIER (ACP).
FT   DOMAIN     2209   2511       THIOESTERASE.
FT   MOD_RES       1      1       ACETYLATION.
FT   ACT_SITE    160    160       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    579    579       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   ACT_SITE    877    877       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   NP_BIND    1674   1691       NADP (ER).
FT   BINDING    1707   1707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1888   1903       NADP (KR).
FT   BINDING    2157   2157       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2308   2308       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2481   2481       THIOESTERASE (BY SIMILARITY).
FT   VARSPLIC   2348   2348       T -> TQCFSFSLF (IN ISOFORM 1).
FT   CONFLICT     77     78       QL -> PV (IN REF. 2).
FT   CONFLICT    116    116       L -> A (IN REF. 2).
FT   CONFLICT    675    675       R -> S (IN REF. 2).
FT   CONFLICT   1169   1169       K -> N (IN REF. 2).
FT   CONFLICT   1178   1178       A -> T (IN REF. 2).
FT   CONFLICT   1191   1191       R -> H (IN REF. 2).
FT   CONFLICT   1198   1198       P -> L (IN REF. 2).
FT   CONFLICT   1286   1287       DN -> ND (IN REF. 2).
FT   CONFLICT   1372   1372       K -> E (IN REF. 2).
FT   CONFLICT   1533   1533       C -> Y (IN REF. 2).
FT   CONFLICT   1577   1577       W -> R (IN REF. 2).
FT   CONFLICT   1685   1696       QAAIAIALSMGC -> ASSHCHRLEHGLA
FT                                (IN REF. 2; AAA48767).
FT   CONFLICT   1732   1732       Q -> E (IN REF. 2).
FT   CONFLICT   1745   1745       S -> N (IN REF. 2).
SQ   SEQUENCE   2511 AA;  274648 MW;  622039DAC8315D3F CRC64;

  Query Match             20.1%;  Score 191.5;  DB 1;  Length 2511;
  Best Local Similarity   29.7%;  Pred. No. 1.1e-08;
  Matches   52;  Conservative   31;  Mismatches   89;  Indels    3;  Gaps    3;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 63
              :|||| |     |   |   :  | |:||||:|  | :      : :  || : ::| |
Db       1890 IITGGLGGFGLELAQWL-IERGAQKLVLTSRSGIRTGYQAKCVREWKALGIQVLVSTSDV 1948

Qy         64 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAK-AHSAHLLHQL 122
                : | |:        |:  : : | :| || : | ||     | : | : : ||
Db       1949 GTLEGTQLLIEEALKLGPVGGIFNLAVVLKDAMIENQTPELFWEVNKPKYSGTLHLDWVT 2008

Qy        123 TQHTP-LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 176
               :  | |  ||::|| :   |  ||:||  ||: :: :   ||   ||  :: ||
Db       2009 RKKCPDLDYFVVFSSVSCGRGNAGQSNYGFANSAMERICEQRHHDGLPGLAVQWG 2063


RESULT 13
FAS_HUMAN
ID   FAS_HUMAN      STANDARD;      PRT;  2504 AA.
AC   P49327;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96004605; PubMed=7567999;
RA   Jayakumar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M.,
RA   Abu-Elheiga L., Chirala S.S., Wakil S.J.;
RT   "Human fatty acid synthase: properties and molecular cloning.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699(1995).
RN   [2]
RP   SEQUENCE OF 753-758 AND 1285-1297.
RX   MEDLINE=94294385; PubMed=8022791;
RA   Kuhajda F.P., Jenner K., Wood F.D., Hennigar R.A., Jacobs L.B.,
RA   Dick J.D., Pasternack G.R.;
RT   "Fatty acid synthesis: a potential selective target for
RT   antineoplastic therapy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383(1994).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND
CC       LIVER.
CC   -!- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY
CC       MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTETHEINE CONTENT OF
CC       THE PROTEIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U26644; AAC50259.1; -.
DR   Genew; HGNC:3594; FASN.
DR   MIM; 600212; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      428    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1630   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2504       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    580    580       MALONYLTRANSFERASE (BY SIMILARITY).
FT   ACT_SITE    876    876       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   NP_BIND    1666   1683       NADP (ER).
FT   BINDING    1699   1699       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1880   1895       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2504 AA;  273100 MW;  8AAF9809B2338DFA CRC64;

  Query Match             19.2%;  Score 182.5;  DB 1;  Length 2504;
  Best Local Similarity   27.4%;  Pred. No. 6.7e-08;
  Matches   48;  Conservative   36;  Mismatches   88;  Indels    3;  Gaps    2;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 63
              :| || |    |   |   :  | |:||||:|  | :      : :::|: : ::| :
Db       1882 IIAGGLGGFGLELAQWL-IQRGVQKLVLTSRSGIRTGYQAKQVRRWRRQGLQVQVSTSNI 1940

Qy         64 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLT 123
              |: :  : |:       |: | : | :| |  | | ||    :| : |       | ::|
Db       1941 SSLEGARGLIAEAAQLGPVGGVFNLAVVLRDGLLENQTPEFFQDVCKPKYSGTLNLDRVT 2000

Qy        124 QHT--PLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 176
                 :    | ||::|| :   |  ||:||  ||: :: :   |    ||  :: ||
Db       2001 REACPELDYFVVFSSVSCGRGNAGQSNYGFANSAMERICEKRRHEGLPGLAVQWG 2055


RESULT 14
FAS_MOUSE
ID   FAS_MOUSE      STANDARD;      PRT;   838 AA.
AC   P19096;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]
DE   (Fragment).
GN   FASN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Liver;
RX   MEDLINE=89149781; PubMed=2920037;
RA   Paulauskis J.D., Sul H.S.;
RT   "Structure of mouse fatty acid synthase mRNA. Identification of the
RT   two NADPH binding sites.";
RL   Biochem. Biophys. Res. Commun. 158:690-695(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X13135; CAA31525.1; -.
DR   PIR; A32262; A32262.
DR   MGD; MGI:95485; Fasn.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; PARTIAL.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP.
FT   NON_TER       1      1
FT   DOMAIN      451    507       ACYL CARRIER (ACP).
FT   DOMAIN      535    838       THIOESTERASE.
FT   BINDING     484    484       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    635    635       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    808    808       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   838 AA;  91212 MW;  32C40EE5CA5AB841 CRC64;

  Query Match             18.8%;  Score 179.5;  DB 1;  Length 838;
  Best Local Similarity   29.5%;  Pred. No. 3.2e-08;
  Matches   52;  Conservative   34;  Mismatches   85;  Indels    5;  Gaps    4;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPH-AQHLTTQLQQKGIHLTITTCD 62
              :|||| |    |   |   :  | |:||||:|  | : |:|:  : :::|| : ::| :
Db        215 IITGGLGGFGLELARWLVL-RGAQRLVLTSRSGIRTGYQAKHI-REWRRQGIQVLVSTSN 272

Qy         63 TSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQL 122
```

```
                |: :  : |:       |:  | :    | || | | ||    :| : | :    | :
Db        273 VSSLEGARALIAEATKLGPVGGVFNLGHGLRDAMLENQTPELFQDVNKPKYNGNLNLDRA 332

Qy        123 TQHT--PLTAPVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 176
              |:     | || :|| :   |  || ||  ||: :: :   |    ||  :: ||
Db        333 TREACPELDYFVAFSSVSCGRGNAGQTNYCFANSTMERICEQRRHDVLPGLAVQWG 388


RESULT 15
PKSM_BACSU
ID   PKSM_BACSU     STANDARD;      PRT;  4273 AA.
AC   P40872; O31781;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksM.
GN   PKSM OR PKSY.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE OF 1-1763 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Tognoni A., Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99113; CAB13603.1; -.
DR   EMBL; Z35133; CAA84505.1; -.
DR   SubtiList; BG10931; pksM.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001601; Methyltransf.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      295    364       ACYL CARRIER (ACP) 1.
FT   DOMAIN      396    834       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     2190   2258       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2322   2737       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     3532   3947       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3410   3483       ACYL CARRIER (ACP) 3.
FT   DOMAIN     4140   4209       ACYL CARRIER (ACP) 4.
FT   BINDING     327    327       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2222   2222       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE   2476   2476       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING    3446   3446       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE   3690   3690       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING    4172   4172       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    103    103       E -> V (IN REF. 2).
FT   CONFLICT    276    276       Q -> E (IN REF. 2).
FT   CONFLICT    289    289       T -> S (IN REF. 2).
SQ   SEQUENCE   4273 AA;  477459 MW;  3BBFCF1A250AEB5A CRC64;

  Query Match             15.3%;  Score 146;  DB 1;  Length 4273;
  Best Local Similarity   26.7%;  Pred. No. 0.0002;
  Matches   48;  Conservative   27;  Mismatches  101;  Indels    4;  Gaps    4;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||| | |      ||  :    :|:|| |: |      |    :|:  |
Db       3145 GTYLITGGVGGLGYLFAKHLAKNY-AANLILTGRS-PFNDEKQKQIKELKDLGGERMYAE 3202

Qy         61 CDTSNPDQLQQLLNTIPPQH-PLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
               | |:|  :   :     :   :  ||| |||  |: : :        ::  | :    |
Db       3203 ADVSDPIAMGDCVKRGKDRFGAINGVIHAAGIESDSAIFDKKIESFQRIIEPKINGTIAL 3262

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPAT-SIAWGTW 178
               :  ::  |     :||::|  |  |  :||  | :  | | :|:  |   |  | | |
Db       3263 DEWLKNEDLDFMCYFSSSSAVLGDFGCCDYAIGNRFQMAYAQYRNELHNGKTFVINWPVW 3322


Search completed: June 17, 2003, 13:02:35
Job time : 2.94115 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 3.97473 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_1715_1892
Perfect score: 953
Sequence: 1 GTTLITGGTGALATHLTHHL.....LAHHRHTHHLPATSIAWGTW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	452	47.4	6260	2	T30228	polyketide synthas
2	449	47.1	8563	2	T30226	polyketide synthas
3	440	46.2	10223	2	T30225	polyketide synthas
4	439	46.1	4735	2	T17463	rifamycin polyketi
5	431	45.2	4151	2	G70944	probable polyketid
6	429.5	45.1	6420	2	T30283	polyketide synthas
7	428.5	45.0	4613	2	T17409	polyketide synthas
8	421.5	44.2	2100	2	T03223	probable polyketid
9	419.5	44.0	7576	2	T17428	FK506 polyketide s
10	419	44.0	2124	2	T28658	polyketide synthas
11	415.5	43.6	2723	2	T03221	probable polyketid
12	411	43.1	3739	2	T17410	polyketide synthas
13	409.5	43.0	502	2	A70985	probable polyketid
14	400.5	42.0	1937	2	T03224	probable polyketid
15	392.5	41.2	1562	2	T17411	polyketide synthas
16	392.5	41.2	1762	2	T03222	probable polyketid
17	391	41.0	2126	2	H70621	probable polyketid
18	389.5	40.9	3491	2	T43231	probable 6-deoxyer
19	372.5	39.1	2103	2	G86925	probable polyketid
20	371.5	39.0	1616	2	G70668	polyketide synthas
21	367.5	38.6	3573	2	S23070	erythronolide synt
22	364.5	38.2	2518	2	AI2140	polyketide synthas
23	351	36.8	3413	2	T17467	rifamycin polyketi
24	349.5	36.7	2478	2	AH2140	polyketide synthas
25	343	36.0	5069	2	T17464	rifamycin polyketi
26	334.5	35.1	1728	2	T17466	rifamycin polyketi
27	326	34.2	3172	2	S22012	erythronolide synt
28	312	32.7	3178	2	S13595	6-deoxyerythronoli
29	304.5	32.0	1763	2	T17465	rifamycin polyketi
30	303	31.8	1570	2	AC2012	hypothetical prote
31	300.5	31.5	1582	2	E70876	probable polyketid
32	295	31.0	3519	2	S43048	polyketide synthas
33	294.5	30.9	2126	2	E70522	probable polyketid
34	286	30.0	1198	2	T28678	polyketide synthas
35	282.5	29.6	2108	2	H70819	probable polyketid
36	278.5	29.2	2111	2	A70668	mycocerosate synth
37	271.5	28.5	2110	2	B44110	mycocerosate synth
38	270.5	28.4	2116	2	C86926	probable mycoceros
39	247.5	26.0	2118	2	S72705	mycocerosate synth
40	233	24.4	958	2	S73012	polyketide synthas
41	233	24.4	1871	2	A87204	polyketide synthas
42	231.5	24.3	2188	2	A70984	probable polyketid
43	223	23.4	4930	2	E69679	polyketide synthet
44	220.5	23.1	1795	2	D97312	hypothetical prote
45	218.5	22.9	3161	2	T30342	protein HMWP1 - Ye

ALIGNMENTS

RESULT 1
T30228
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30228
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30228
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6260 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987102; PIDN:CAA60462.1
C;Genetics:
A;Gene: rapC
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1536-1607/Domain: acyl carrier protein homology <ACP1>
F;1651-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3123-3194/Domain: acyl carrier protein homology <ACP2>
F;3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3725-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5114-5185/Domain: acyl carrier protein homology <ACP3>
F;5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6135-6206/Domain: acyl carrier protein homology <ACP4>

Query Match 47.4%; Score 452; DB 2; Length 6260;
Best Local Similarity 52.8%; Pred. No. 6.2e-31;
Matches 94; Conservative 19; Mismatches 61; Indels 4; Gaps 1;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||:| ||  |  ||      :|||| ||| |     | |  :| : | |:   |
Db       4840 GTILITGGSGVLAGILARHLAAEHGARHLLLLSRTTPD----QALIKELAELGAHVDTAT 4895

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || |:   | ::|  : |:|||| |||||| |||  : :||  ||: ||| ||  |  ||
Db       4896 CDVSDRAGLARVLAGVSPEHPLTAVIHTAGALDDGVVESLTTQQLDTVLRPKADGAWHLH 4955

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||:| | |||:|||||   |  || |||||||:|||||  |    ||| ::||| |
Db       4956 ELTQNTDLAAFVMYSSAAGVIGGAGQGNYAAANAFLDALAEQRRAEGLPALALAWGLW 5013
```

RESULT 2
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000

```
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             47.1%;  Score 449;  DB 2;  Length 8563;
  Best Local Similarity   52.2%;  Pred. No. 1.7e-30;
  Matches   93;  Conservative   21;  Mismatches   60;  Indels    4;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||:| ||  |  ||     :|||| ||| |     | |  :| : | |:   |
Db       2931 GTILITGGSGVLAGILARHLAAEHGARHLLLLSRTTPD----QALIKELAELGAHVDTAT 2986

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || |:   | ::|  : |:|||| |||||| |||  : :||  ||: ||| ||  |  ||
Db       2987 CDVSDRAGLARVLAGVSPEHPLTAVIHTAGALDDGVVESLTTQQLDTVLRPKADGAWHLH 3046

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||:| | |||:|||||   |: || |||||||::|||  |    ||| ::||| |
Db       3047 ELTQNTDLAAFVMYSSAAGVLGSAGQGNYAAANAFVDALAEQRRGEGLPALAVAWGLW 3104


RESULT 3
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match             46.2%;  Score 440;  DB 2;  Length 10223;
  Best Local Similarity   50.6%;  Pred. No. 1.3e-29;
  Matches   90;  Conservative   23;  Mismatches   61;  Indels    4;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||:| ||     || | :  :|||| ||: |     : |  || : |  :
Db       8145 GTVLITGGSGVLAGIAVRHLVTERGVRHLLLLSRSAPD----EALINQLGELGARVETAA 8200

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || |:   | |:|  : |:|||| ||||||:|||  : :||  :|: ||| ||  |  ||
Db       8201 CDVSDRAALAQVLAGVSPEHPLTAVIHTAGVLDDGVVESLTAQRLDAVLRPKADGAWNLH 8260

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | |||:|||||   |: || |||||||:|||||  ||   ||| ::||| |
Db       8261 ELTRDADLAAFVMYSSAAGVLGSGGQGNYAAANAFLDALAEQRHAEGLPALAVAWGLW 8318


RESULT 4
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

  Query Match             46.1%;  Score 439;  DB 2;  Length 4735;
  Best Local Similarity   51.7%;  Pred. No. 6.2e-30;
  Matches   92;  Conservative   18;  Mismatches   64;  Indels    4;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |      || |    :||:| || |   . | |  :|   | : |
Db       1829 GTVLITGGTGTLGALTARHLVTAHGVRHLVLVSRRG----EAPELQEELTALGASVAIAA 1884

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::  ||: :|  || :|||| ||||||:|||  :| ||| :|  | | |  :| ||
Db       1885 CDVADRAQLEAVLRAIPAEHPLTAVIHTAGVLDDGVVTELTPDRLATVRRPKVDAARLLD 1944

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:||||   | |||| |||||| |||||  |::  ||| ||||| |
Db       1945 ELTREADLAAFVLFSSAAGVLGNPGQAGYAAANAELDALARQRNSLDLPAVSIAWGYW 2002


RESULT 5
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
```

```
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

  Query Match             45.2%;  Score 431;  DB 2;  Length 4151;
  Best Local Similarity   46.6%;  Pred. No. 2.7e-29;
  Matches   83;  Conservative   32;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||   : :  |:      ::|:| || ||  | |  | :|   |  : :
Db       1680 GTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVAELAAAGAQVQVVA 1739

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::   | :::  || |||: ||||| |||| : :||| ::: |||:|  :|  ||
Db       1740 CDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRVDVVLRSKVDAAWHLH 1799

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   ::|||::|| |   |: |||||||||::||||| ||  | ||| |: || |
Db       1800 ELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALAAHRRAHGLPAISLGWGLW 1857


RESULT 6
T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30283
R;Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A;Title: Structural organization of a multifunctional polyketide synthase involved in th
A;Reference number: Z20806; MUID:97217427; PMID:9063448
A;Accession: T30283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6420 <MOT>
A;Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C;Genetics:
A;Note: fkbA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3391-3462/Domain: acyl carrier protein homology <ACP1>
F;3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5307-5378/Domain: acyl carrier protein homology <ACP2>
F;5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6293-6364/Domain: acyl carrier protein homology <ACP3>

  Query Match             45.1%;  Score 429.5;  DB 2;  Length 6420;
  Best Local Similarity   54.3%;  Pred. No. 6.2e-29;
  Matches   95;  Conservative   15;  Mismatches   46;  Indels   19;  Gaps    4;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 63
              |||||:| ||  |  ||  : |  :|| :   | ||                | |:   ||
Db       3140 LITGGSGTLAGILARHL--NHPHTYLLSRTPPPPTTP-----------GTHI---PCDL 3182

Qy         64 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLT 123
              ::| |: | |  ||   |||: ||| |||||||||||  |   |: || :|  ||  |
Db       3183 TDPTQITQALTHIP--QPLTGIFHTAATLDDATLTNLTPQHLTTTLQPKADAAWHLHHHT 3240

Qy        124 QHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |: ||| |||||||||| |:|||||||||:|||| ||||   |||:|||| |
Db       3241 QNQPLTHFVLYSSAAATLGSPGQANYAAANAFLDALATHRHTQGQPATTIAWGMW 3295


RESULT 7
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

  Query Match             45.0%;  Score 428.5;  DB 2;  Length 4613;
  Best Local Similarity   48.9%;  Pred. No. 5.1e-29;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2208 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2266

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2267 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2326

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db       2327 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2384


RESULT 8
T03223
probable polyketide synthase module 3 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03223
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2100 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38063.1; PID:g2624950
A;Experimental source: ATCC 29253
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I h
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
```

```
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;566-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1952-2027/Domain: acyl carrier protein homology <ACP>

  Query Match             44.2%;  Score 421.5;  DB 2;  Length 2100;
  Best Local Similarity   49.4%;  Pred. No. 7.9e-29;
  Matches   88;  Conservative   21;  Mismatches   66;  Indels    3;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| | : : ||| |    : |:|||| ||: | |  || |||| |  :   :
Db       1677 GTILITGGTGNLGSLIAHHLVTTHGARDLILTSRQGPNAPGATELTNQLQQHGARIRTIS 1736

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::   |  |:: |    ||| |:|||| | | |: :| |  |   |  ||: |  ||
Db       1737 CDLTDRTALTTLIDEI---GPLTGVVHTAGALADTTIDHLNPDTLTTTLTPKANPAWWLH 1793

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||   |  ||::|| |   |  |||||||||::|||| ||    || ||:||| |
Db       1794 ELTQDHDLALFVVFSSVAGVLGNAGQANYAAANSFLDALITHRRRQGLPGTSLAWGWW 1851


RESULT 9
T17428
FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: T17428
R;Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressan
A;Reference number: Z18779; MUID:98451508; PMID:9780228
A;Accession: T17428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7576 <MOT>
A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A;Experimental source: strain MA6548
C;Genetics:
A;Gene: fkbB
C;Function:
A;Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-500/Domain: acetate-CoA ligase homology <ACL>
F;1095-1166/Domain: acyl carrier protein homology <ACP1>
F;1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;2680-2751/Domain: acyl carrier protein homology <ACP2>
F;2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;4320-4391/Domain: acyl carrier protein homology <ACP3>
F;4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5903-5974/Domain: acyl carrier protein homology <ACP4>
F;6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;7450-7521/Domain: acyl carrier protein homology <ACP5>

  Query Match             44.0%;  Score 419.5;  DB 2;  Length 7576;
  Best Local Similarity   53.1%;  Pred. No. 5.9e-28;
  Matches   93;  Conservative   14;  Mismatches   49;  Indels   19;  Gaps    3;

Qy          4 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 63
              |||||:| ||  |  ||  : |  :||  :   | ||          .       |   ||
Db       2454 LITGGSGTLAGILARHL--NHPLTYLLSRTPPPPTTPR--------------THIPCDL 2496

Qy         64 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLT 123
              ::| |: | |  ||   ||| : |||   ||||||||||  |   |: || :|  ||  |
Db       2497 TDPTQITQALTHIP--QPLTGIFHTAATHDDATLTNLTPQHLTTTLQPKADAAWHLHHHT 2554

Qy        124 QHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |: ||| |||||||||| |:|||||||||||:||||| ||||   |||:|||| |
Db       2555 QNQPLTHFVLYSSAAATLGSPGQANYAAANAFLDALATHRHTQGQPATTIAWGMW 2609


RESULT 10
T28658
polyketide synthase - Sorangium cellulosum (fragment)
C;Species: Sorangium cellulosum
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C;Accession: T28658
R;Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J.J.; Ligon, J.M.
J. Bacteriol. 177, 3673-3679, 1995
A;Title: A Sorangium cellulosum (myxobacterium) gene cluster for the biosynthesis of th
inomycetes.
A;Reference number: Z20499; MUID:95325306; PMID:7601830
A;Accession: T28658
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2124 <SCH>
A;Cross-references: EMBL:U24241; NID:g868150; PID:g868151; PIDN:AAA79984.1
A;Experimental source: strain So ce26
C;Function:
A;Description: involved in soraphen A biosynthesis
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;335-406/Domain: acyl carrier protein homology <ACP>
F;370/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             44.0%;  Score 419;  DB 2;  Length 2124;
  Best Local Similarity   45.5%;  Pred. No. 1.3e-28;
  Matches   81;  Conservative   29;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |   :   |  :   :|||||| |   | | | ::|:  |  :|:
Db       1819 GTVLITGGTGTLGALVARRLVVNHDAKHLLLTSRQGASAPGADVLRSELEALGASVTLAA 1878

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  |: ||: ||  ||:  |:| | :||   |  ::  ::: |  |  :|  ||
Db       1879 CDVADPRALKDLLDNIPSAHPVAAVVHAASVLDGDLLGAMSLERIDRVFAPKIDAAWHLH 1938

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||||  || ||:|:|| |   |: | :|||||:|:|||||||    |||:|:||  |
Db       1939 QLTQDKPLAAFILFSSVAGVLGSSGHSNYAAASAFLDALAHHRRAQGLPASSLAWSHW 1996


RESULT 11
T03221
probable polyketide synthase module 1 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 17-Nov-2000
C;Accession: T03221
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03221
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2723 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38061.1; PID:g2624948
A;Experimental source: ATCC 29253
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-504/Domain: acetate-CoA ligase homology <ACL>
F;881-949/Domain: acyl carrier protein homology <ACP1>
F;996-1397/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;1509-1794/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;2568-2639/Domain: acyl carrier protein homology <ACP2>

  Query Match             43.6%;  Score 415.5;  DB 2;  Length 2723;
  Best Local Similarity   47.8%;  Pred. No. 3.7e-28;
  Matches   85;  Conservative   28;  Mismatches   62;  Indels    3;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
```

```
                 || |||||||:| : :  || |    : |:| || ||  | |  |  :||| |  : : :
Db       2292 GTVLITGGTGSLGSLIARHLVTEHGVRDLVLVSRQGPDAPGATELIAELQQLGAQVRVRS 2351

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               |||::   |  |:: :    |   |:|||| | |||: ||    |   : ||||:|  ||
Db       2352 CDTTDRAALAALVDEV---GPFAGVVHTAGALADATVDNLDADALATTIAAKAHAAWWLH 2408

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               : ||   || |:::|| |:  |: |||||||||::|||| || |  || ||:|||:|
Db       2409 EHTQDHGLTLFMVFSSLASVLGSAGQANYAAANSFLDALATHRRTLGLPGTSLAWGSW 2466


RESULT 12
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

  Query Match             43.1%;  Score 411;  DB 2;  Length 3739;
  Best Local Similarity   47.5%;  Pred. No. 1.4e-27;
  Matches   86;  Conservative   27;  Mismatches   64;  Indels    4;  Gaps    2;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||||    :   |:       : ||| || |    | |  |  :|:  |  :::
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || :: : |  :|::|| :|||| |:||||:| | || ::|    : :||| |   :| ||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 QLTQHTP---LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGT 177
              :||  ||   | |||::||||| ||  ||  |||||| ||||| | |  ||| |: ||
Db       3397 ELTS-TPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGL 3455

Qy        178 W 178
               |
Db       3456 W 3456


RESULT 13
A70985
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-May-2000
C;Accession: A70985
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70985
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-502 <COL>
A;Cross-references: GB:Z95617; GB:AL123456; NID:g3242249; PIDN:CAB09099.1; PID:e317263;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks17
C;Superfamily: acyl carrier protein homology; short-chain alcohol dehydrogenase homolog
C;Keywords: carrier protein
F;64-242/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;340-414/Domain: acyl carrier protein homology <ACP2>

  Query Match             43.0%;  Score 409.5;  DB 2;  Length 502;
  Best Local Similarity   46.1%;  Pred. No. 1.5e-28;
  Matches   82;  Conservative   28;  Mismatches   65;  Indels    3;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||   : :  || :    : ::| || | ||      |  :|   |  : : :
Db         64 GTVLITGGTGMAGSAVARHLVSRYGVRQVVLASRAGEHTESVAALVDELGSAGARVQVVS 123

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || :: | :  |   :  |  || | | ||:|||| :| ||| ::: |||||   |  ||
Db        124 CDVADRDAVAGL---VASQPDLTAVFHAAGVLDDAVITGLTPERVDKVLRAKVDGAWNLH 180

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:|  ::||||:|| |   |||||||||||||:|| || :| :  | | |:||| |
Db        181 ELTRHLDVSAFVLFSSMAGIVGAPGQANYAAANAFLDGLAAYRRSRGLAALSVAWGLW 238


RESULT 14
T03224
probable polyketide synthase module 4 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03224
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03224
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1937 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38064.1; PID:g2624951
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;549-827/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

  Query Match             42.0%;  Score 400.5;  DB 2;  Length 1937;
  Best Local Similarity   47.8%;  Pred. No. 5.2e-27;
  Matches   85;  Conservative   23;  Mismatches   67;  Indels    3;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||:| : :  || |    : |:| || ||    |  |  :||| |  :   :
Db       1319 GTILITGGTGSLGSLVARHLVTEHGVRDLVLVSRQGPDADSATELAAELQQHGARVRTLS 1378

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || ::   |  |:: |    ||| |:|||| | |||: :|    |       ||::|  ||
Db       1379 CDLTDRTALATLIDEI---GPLTGVVHTAGALADATVEHLDAEALATTFAPKANTAWWLH 1435

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||   || | |:|| |   |:|||||||||| :||||| ||    || ||:|||:|
Db       1436 ELTQDHDLTLFALFSSVAGILGSPGQANYAAANTFLDALATHRRGLGLPGTSLAWGSW 1493


RESULT 15
T17411
polyketide synthase III - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17411
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
```

A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17411
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1562 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C;Genetics:
A;Gene: pikAIII
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1403-1474/Domain: acyl carrier protein homology <ACP>

Query Match 41.2%; Score 392.5; DB 2; Length 1562;
Best Local Similarity 46.6%; Pred. No. 2e-26;
Matches 83; Conservative 21; Mismatches 73; Indels 1; Gaps 1;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||  :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       1116 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 1174

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:||||      |    |  :  :| ||   | :|
Db       1175 CDVADPHAMRTLLDAIPAETPLTAVVHTAGAPGGDPLDVTGPEDIARILGAKTSGAEVLD 1234

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               | : ||| ||||||| |  :|:  |  ||||||:|||||  |      |||:||| |
Db       1235 DLLRGTPLDAFVLYSSNAGVWGSGSQGVYAAANAHLDALAARRRARGETATSVAWGLW 1292
```

Search completed: June 17, 2003, 13:12:21
Job time : 3.97473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 5.54613 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_1715_1892
Perfect score: 953
Sequence: 1 GTTLITGGTGALATHLTHHL.....LAHHRHTHHLPATSIAWGTW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	428.5	45.0	4551	9	US-09-793-708-1	Sequence 1, Appli
2	428.5	45.0	4613	9	US-09-860-846-31	Sequence 31, Appl
3	428.5	45.0	4613	9	US-09-988-384B-31	Sequence 31, Appl
4	428.5	45.0	4613	9	US-09-836-821-31	Sequence 31, Appl
5	428.5	45.0	4613	10	US-09-861-289-31	Sequence 31, Appl
6	428.5	45.0	11877	9	US-09-860-846-6	Sequence 6, Appli
7	428.5	45.0	11877	9	US-09-836-821-6	Sequence 6, Appli
8	428.5	45.0	11877	10	US-09-861-289-6	Sequence 6, Appli
9	428.5	45.0	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	412	43.2	5215	9	US-09-860-846-2	Sequence 2, Appli
11	412	43.2	5215	9	US-09-988-384B-2	Sequence 2, Appli
12	412	43.2	5215	9	US-09-836-821-2	Sequence 2, Appli
13	412	43.2	5215	10	US-09-861-289-2	Sequence 2, Appli
14	411	43.1	3739	9	US-09-860-846-33	Sequence 33, Appl
15	411	43.1	3739	9	US-09-988-384B-33	Sequence 33, Appl
16	411	43.1	3739	9	US-09-836-821-33	Sequence 33, Appl
17	411	43.1	3739	9	US-09-793-708-2	Sequence 2, Appli
18	411	43.1	3739	10	US-09-861-289-33	Sequence 33, Appl
19	400	42.0	4150	9	US-09-808-880-2	Sequence 2, Appli
20	392.5	41.2	1562	9	US-09-860-846-35	Sequence 35, Appl
21	392.5	41.2	1562	9	US-09-988-384B-35	Sequence 35, Appl
22	392.5	41.2	1562	9	US-09-836-821-35	Sequence 35, Appl
23	392.5	41.2	1562	9	US-09-793-708-3	Sequence 3, Appli
24	392.5	41.2	1562	10	US-09-861-289-35	Sequence 35, Appl
25	371.5	39.0	1616	9	US-09-712-363-262	Sequence 262, App
26	367.5	38.6	3816	9	US-09-808-880-3	Sequence 3, Appli
27	362.5	38.0	7257	9	US-10-014-717-5	Sequence 5, Appli
28	349	36.6	184	9	US-09-957-483-33	Sequence 33, Appl
29	345	36.2	2439	9	US-10-014-717-7	Sequence 7, Appli
30	337	35.4	1832	9	US-10-014-717-4	Sequence 4, Appli
31	335	35.2	184	9	US-09-957-483-35	Sequence 35, Appl
32	328.5	34.5	3798	9	US-10-014-717-6	Sequence 6, Appli
33	295	31.0	3519	9	US-09-808-880-4	Sequence 4, Appli
34	263.5	27.6	2563	9	US-09-836-705-46	Sequence 46, Appl
35	210	22.0	1827	9	US-09-712-363-261	Sequence 261, App
36	199	20.9	3032	9	US-09-836-705-44	Sequence 44, Appl
37	181.5	19.0	2509	9	US-10-237-271-1	Sequence 1, Appli
38	117	12.3	2756	9	US-10-331-061-7	Sequence 7, Appli
39	100.5	10.5	711	9	US-10-124-800-18	Sequence 18, Appl
40	100.5	10.5	1481	9	US-10-331-061-70	Sequence 70, Appl
41	100.5	10.5	2910	9	US-10-124-800-2	Sequence 2, Appli
42	98.5	10.3	244	10	US-09-815-242-10126	Sequence 10126, A
43	97.5	10.2	243	10	US-09-815-242-13360	Sequence 13360, A
44	97.5	10.2	243	10	US-09-815-242-13581	Sequence 13581, A
45	97.5	10.2	244	9	US-09-849-092-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-793-708-1

Query Match 45.0%; Score 428.5; DB 9; Length 4551;
Best Local Similarity 48.9%; Pred. No. 1.1e-31;
Matches 87; Conservative 22; Mismatches 68; Indels 1; Gaps 1;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2146 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2204
```

```
Qy          61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    |||| | :|
Db        2205 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2264

Qy         121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db        2265 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2322


RESULT 2
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-31

  Query Match             45.0%;  Score 428.5;  DB 9;  Length 4613;
  Best Local Similarity   48.9%;  Pred. No. 1.1e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy           1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
               || |||||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db        2208 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2266

Qy          61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    |||| | :|
Db        2267 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2326

Qy         121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db        2327 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2384


RESULT 3
US-09-988-384B-31
; Sequence 31, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-31

  Query Match             45.0%;  Score 428.5;  DB 9;  Length 4613;
  Best Local Similarity   48.9%;  Pred. No. 1.1e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy           1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
               || |||||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db        2208 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2266

Qy          61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    |||| | :|
Db        2267 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2326

Qy         121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db        2327 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2384


RESULT 4
US-09-836-821-31
; Sequence 31, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-31

  Query Match             45.0%;  Score 428.5;  DB 9;  Length 4613;
  Best Local Similarity   48.9%;  Pred. No. 1.1e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy           1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
               || |||||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db        2208 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2266

Qy          61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
               || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    |||| | :|
Db        2267 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2326

Qy         121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db        2327 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2384


RESULT 5
US-09-861-289-31
; Sequence 31, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-31

  Query Match             45.0%;  Score 428.5;  DB 10;  Length 4613;
  Best Local Similarity   48.9%;  Pred. No. 1.1e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||  :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2208 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2266

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2267 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2326

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  ||||||||  |       | |:||| |
Db       2327 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2384


RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match             45.0%;  Score 428.5;  DB 9;  Length 11877;
  Best Local Similarity   48.9%;  Pred. No. 3.9e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||  :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2544 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2602

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2603 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2662

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  ||||||||  |       | |:||| |
Db       2663 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2720


RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-6

  Query Match             45.0%;  Score 428.5;  DB 9;  Length 11877;
  Best Local Similarity   48.9%;  Pred. No. 3.9e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||  :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2544 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2602

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2603 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2662

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  ||||||||  |       | |:||| |
Db       2663 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2720


RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             45.0%;  Score 428.5;  DB 10;  Length 11877;
  Best Local Similarity   48.9%;  Pred. No. 3.9e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||  :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2544 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2602

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2603 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2662

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  ||||||||  |       | |:||| |
Db       2663 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2720


RESULT 9
US-09-988-384B-6
```

```
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             45.0%;  Score 428.5;  DB 9;  Length 12199;
  Best Local Similarity   48.9%;  Pred. No. 4.1e-31;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||||| :|     :  |   :|||| ||:|   | |  || :|   |  :||
Db       2866 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2924

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2925 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2984

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db       2985 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 3042


RESULT 10
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-2

  Query Match             43.2%;  Score 412;  DB 9;  Length 5215;
  Best Local Similarity   47.8%;  Pred. No. 5e-30;
  Matches   85;  Conservative   22;  Mismatches   71;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| |   | ||      : ||| || |   | | |   |   | :
Db       2440 GTVLVTGGTGGLGALLARHLVERHGVRRLLLVSRRGADAPGAADLGEDLAGLGAEVAFAA 2499

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
                | :: : | : : |:| :|||| |:||||::||||:  ||| :|: ||| |  :|  ||
Db       2500 ADAADRESLARAIATVPAEHPLTAVVHTAGVVDDATVEALTPERLDAVLRPKVDAAWNLH 2559

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| :   |  |||||||||  ||||| ||    | |||:||| |
Db       2560 ELTKDLRLDAFVLFSSVSGIVGTAGQANYAAANTGLDALAAHRAATGLAATSLAWGLW 2617


RESULT 11
US-09-988-384B-2
; Sequence 2, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-2

  Query Match             43.2%;  Score 412;  DB 9;  Length 5215;
  Best Local Similarity   47.8%;  Pred. No. 5e-30;
  Matches   85;  Conservative   22;  Mismatches   71;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| |   |  |  |      : ||| || |   | | |   |   | :
Db       2440 GTVLVTGGTGGLGALLARHLVERHGVRRLLLVSRRGADAPGAADLGEDLAGLGAEVAFAA 2499

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
                | :: : | : : |:| :|||| |:||||::||||:  ||| :|: ||| |  :|  ||
Db       2500 ADAADRESLARAIATVPAEHPLTAVVHTAGVVDDATVEALTPERLDAVLRPKVDAAWNLH 2559

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| :   |  |||||||||  ||||| ||    | |||:||| |
Db       2560 ELTKDLRLDAFVLFSSVSGIVGTAGQANYAAANTGLDALAAHRAATGLAATSLAWGLW 2617


RESULT 12
US-09-836-821-2
; Sequence 2, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-2

  Query Match             43.2%;  Score 412;  DB 9;  Length 5215;
  Best Local Similarity   47.8%;  Pred. No. 5e-30;
  Matches   85;  Conservative   22;  Mismatches   71;  Indels    0;  Gaps    0;
```

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||  |    |  ||       : ||| || |    | |  |    |   | :
Db       2440 GTVLVTGGTGGLGALLARHLVERHGVRRLLLVSRRGADAPGAADLGEDLAGLGAEVAFAA 2499

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              |.:: : | : : |:| :|||| |:||||::||||:  ||| :|: ||| |  :|  ||
Db       2500 ADAADRESLARAIATVPAEHPLTAVVHTAGVVDDATVEALTPERLDAVLRPKVDAAWNLH 2559

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| :   |  |||||||||  ||||| ||    | |||:||| |
Db       2560 ELTKDLRLDAFVLFSSVSGIVGTAGQANYAAANTGLDALAAHRAATGLAATSLAWGLW 2617


RESULT 13
US-09-861-289-2
; Sequence 2, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-2

  Query Match             43.2%;  Score 412;  DB 10;  Length 5215;
  Best Local Similarity   47.8%;  Pred. No. 5e-30;
  Matches   85;  Conservative   22;  Mismatches   71;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||  |    |  ||       : ||| || |    | |  |    |   | :
Db       2440 GTVLVTGGTGGLGALLARHLVERHGVRRLLLVSRRGADAPGAADLGEDLAGLGAEVAFAA 2499

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              | :: : | : : |:| :|||| |:||||::||||:  ||| :|: ||| |  :|  ||
Db       2500 ADAADRESLARAIATVPAEHPLTAVVHTAGVVDDATVEALTPERLDAVLRPKVDAAWNLH 2559

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| :   |  |||||||||  ||||| ||    | |||:||| |
Db       2560 ELTKDLRLDAFVLFSSVSGIVGTAGQANYAAANTGLDALAAHRAATGLAATSLAWGLW 2617


RESULT 14
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-33

  Query Match             43.1%;  Score 411;  DB 9;  Length 3739;
  Best Local Similarity   47.5%;  Pred. No. 4.1e-30;
  Matches   86;  Conservative   27;  Mismatches   64;  Indels    4;  Gaps    2;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||||   :  |:      : ||| || |    | |  |  :|:  |  :::
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : |  :|::|| :|||| |:||||:| | || ::|   : :||| |  :| ||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 QLTQHTP---LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGT 177
              :||  ||   | |||::|||||  ||  ||  ||||| |||||  | |  ||| |: ||
Db       3397 ELTS-TPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGL 3455

Qy        178 W 178
              |
Db       3456 W 3456


RESULT 15
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match             43.1%;  Score 411;  DB 9;  Length 3739;
  Best Local Similarity   47.5%;  Pred. No. 4.1e-30;
  Matches   86;  Conservative   27;  Mismatches   64;  Indels    4;  Gaps    2;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||||   :  |:      : ||| || |    | |  |  :|:  |  :::
Db       3277 GTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAA 3336

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : |  :|::|| :|||| |:||||:| | || ::|   : :||| |  :| ||
Db       3337 CDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLD 3396

Qy        121 QLTQHTP---LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGT 177
              :||  ||   | |||::|||||  ||  ||  ||||| |||||  | |  ||| |: ||
Db       3397 ELTS-TPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGL 3455

Qy        178 W 178
              |
Db       3456 W 3456


Search completed: June 17, 2003, 13:46:43
Job time : 7.54613 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 3.23524 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_1715_1892
Perfect score: 953
Sequence: 1 GTTLITGGTGALATHLTHHL.....LAHHRHTHHLPATSIAWGTW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	457	48.0	1864	2	US-08-804-227C-3	Sequence 3, Appli
2	439	46.1	4545	2	US-08-804-227C-14	Sequence 14, Appl
3	439	46.1	4550	2	US-08-804-227C-8	Sequence 8, Appli
4	439	46.1	4550	2	US-08-804-198-2	Sequence 2, Appli
5	436	45.8	5588	4	US-09-036-987A-6	Sequence 6, Appli
6	436	45.8	5588	4	US-09-370-700-6	Sequence 6, Appli
7	428.5	45.0	4551	3	US-09-320-878-1	Sequence 1, Appli
8	428.5	45.0	4613	4	US-09-105-537-31	Sequence 31, Appl
9	428.5	45.0	11877	4	US-09-105-537-6	Sequence 6, Appli
10	423	44.4	1996	2	US-08-804-227C-9	Sequence 9, Appli
11	423	44.4	1996	2	US-08-804-198-3	Sequence 3, Appli
12	423	44.4	4472	2	US-08-804-227C-2	Sequence 2, Appli
13	421	44.2	3724	2	US-08-804-227C-10	Sequence 10, Appl
14	421	44.2	3724	2	US-08-804-198-4	Sequence 4, Appli
15	413.5	43.4	6095	4	US-09-144-085-2	Sequence 2, Appli
16	412	43.2	5215	4	US-09-105-537-2	Sequence 2, Appli
17	411	43.1	3729	2	US-08-804-227C-4	Sequence 4, Appli
18	411	43.1	3739	3	US-09-320-878-2	Sequence 2, Appli
19	411	43.1	3739	4	US-09-105-537-33	Sequence 33, Appl
20	400	42.0	4150	4	US-09-428-517-2	Sequence 2, Appli
21	394.5	41.4	5087	4	US-09-144-085-1	Sequence 1, Appli
22	392.5	41.2	1562	3	US-09-320-878-3	Sequence 3, Appli
23	392.5	41.2	1562	4	US-09-105-537-35	Sequence 35, Appl
24	391	41.0	3567	2	US-07-642-734C-4	Sequence 4, Appli
25	391	41.0	3567	3	US-08-439-009A-4	Sequence 4, Appli
26	389.5	40.9	3491	2	US-07-642-734C-2	Sequence 2, Appli
27	389.5	40.9	3491	3	US-08-439-009A-2	Sequence 2, Appli
28	389	40.8	2152	4	US-09-036-987A-3	Sequence 3, Appli
29	389	40.8	2152	4	US-09-370-700-3	Sequence 3, Appli
30	382.5	40.1	4928	4	US-09-036-987A-5	Sequence 5, Appli
31	382.5	40.1	4928	4	US-09-370-700-5	Sequence 5, Appli
32	373.5	39.2	1841	2	US-08-804-227C-6	Sequence 6, Appli
33	367.5	38.6	3816	4	US-09-428-517-3	Sequence 3, Appli
34	362.5	38.0	7257	3	US-09-335-409-5	Sequence 5, Appli
35	362.5	38.0	7257	4	US-09-568-102-5	Sequence 5, Appli
36	362.5	38.0	7257	4	US-09-567-969-5	Sequence 5, Appli
37	362.5	38.0	7257	4	US-09-568-480-5	Sequence 5, Appli
38	362.5	38.0	7257	4	US-09-568-486-5	Sequence 5, Appli
39	362.5	38.0	7257	4	US-09-568-472-5	Sequence 5, Appli
40	362.5	38.0	7257	4	US-09-567-899-5	Sequence 5, Appli
41	355.5	37.3	2595	4	US-09-036-987A-2	Sequence 2, Appli
42	355.5	37.3	2595	4	US-09-370-700-2	Sequence 2, Appli
43	355	37.3	159	3	US-09-010-809-1	Sequence 1, Appli
44	345	36.2	2259	4	US-09-413-814-70	Sequence 70, Appl
45	345	36.2	2439	3	US-09-335-409-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-3
; Sequence 3, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1864 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-3

Query Match 48.0%; Score 457; DB 2; Length 1864;
Best Local Similarity 50.6%; Pred. No. 3.3e-43;
Matches 90; Conservative 22; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||:|      || |    : ||||||:||  | |  |  :|   | |  :
Db       1414 GTVLITGGTGSLGAEAARHLVTRHGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAA 1473

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
```

```
              || ::   |: ||: :|  |||| |:||||:|||  ||  || :|  ||| || :   ||
Db       1474 CDVADRAALRALLDRVPAGHPLTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLH 1533

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||   |:||:|||||   |: ||: |||||||||: |  | :  ||| |: || |
Db       1534 ELTQGHALSAFILYSSAAGVLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPW 1591


RESULT 2
US-08-804-227C-14
; Sequence 14, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-14

  Query Match             46.1%;  Score 439;  DB 2;  Length 4545;
  Best Local Similarity   49.4%;  Pred. No. 1.4e-40;
  Matches   88;  Conservative   23;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||   :  ||      :|:|| || ||  | |  |  :|   |:  |
Db       4080 GTVLITGGTGALGRRVASHLARRYGVRHMLLVSRRGPDAPEAGPLERELAGLGVTATFLA 4139

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : ::: :  :|  |||| |:||||:|||  || ||  :|: ||| || :   ||
Db       4140 CDLTDIEAVRKAVAAVPSDHPLTGVVHTAGVLDDGALTGLTRQRLDTVLRPKADAVRNLH 4199

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |   || ||||:|:||   | ||||:||||| |||||  |    ||| |:||| |
Db       4200 EATLDRPLRAFVLFSAAAGLLGRPGQASYAAANAVLDALAGARRAAGLPAVSLAWGLW 4257


RESULT 3
US-08-804-227C-8
; Sequence 8, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4550 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-8

  Query Match             46.1%;  Score 439;  DB 2;  Length 4550;
  Best Local Similarity   49.4%;  Pred. No. 1.4e-40;
  Matches   88;  Conservative   23;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||   :  ||      :|:|| || ||  | |  |  :|   |:  |
Db       4085 GTVLITGGTGALGRRVASHLARRYGVRHMLLVSRRGPDAPEAGPLERELAGLGVTATFLA 4144

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : ::: :  :|  |||| |:||||:|||  || ||  :|: ||| || :   ||
Db       4145 CDLTDIEAVRKAVAAVPSDHPLTGVVHTAGVLDDGALTGLTRQRLDTVLRPKADAVRNLH 4204

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |   || ||||:|:||   | ||||:||||| |||||  |    ||| |:||| |
Db       4205 EATLDRPLRAFVLFSAAAGLLGRPGQASYAAANAVLDALAGARRAAGLPAVSLAWGLW 4262


RESULT 4
US-08-804-198-2
; Sequence 2, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
```

```
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4550 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-2

  Query Match             46.1%;  Score 439;  DB 2;  Length 4550;
  Best Local Similarity   49.4%;  Pred. No. 1.4e-40;
  Matches   88;  Conservative   23;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||    :  ||      :|:|| || ||  | |  | :|   |:  |
Db       4085 GTVLITGGTGALGRRVASHLARRYGVRHMLLVSRRGPDAPEAGPLERELAGLGVTATFLA 4144

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : ::: :  :|  |||| |:||||:|||  || ||  :|: ||| || :   ||
Db       4145 CDLTDIEAVRKAVAAVPSDHPLTGVVHTAGVLDDGALTGLTRQRLDTVLRPKADAVRNLH 4204

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |   || ||||:|:||   | ||||:|||||| |||||  |    ||| |:||| |
Db       4205 EATLDRPLRAFVLFSAAAGLLGRPGQASYAAANAVLDALAGARRAAGLPAVSLAWGLW 4262


RESULT 5
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match             45.8%;  Score 436;  DB 4;  Length 5588;
  Best Local Similarity   47.2%;  Pred. No. 4.2e-40;
  Matches   84;  Conservative   31;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| |    |  || :    :|||| || |   | | |  :|:| |  : :
Db       4933 GTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGADELVAELEQAGADVAVVA 4992

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||::: | | :|: ::| ::||  |:| ||:|||  | :::| :|: ||| |  :|  ||
Db       4993 CDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSPERLDAVLRPKVDAAWYLH 5052

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   |:||||:|| |  ||  ||:|||| ||:|||||| |    ||| |:| | |
Db       5053 ELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHCRQAQGLPALSLASGLW 5110


RESULT 6
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match             45.8%;  Score 436;  DB 4;  Length 5588;
  Best Local Similarity   47.2%;  Pred. No. 4.2e-40;
  Matches   84;  Conservative   31;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| |    |  || :    :|||| || |   | | |  :|:| |  : :
Db       4933 GTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGADELVAELEQAGADVAVVA 4992

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||::: | | :|: ::| ::||  |:| ||:|||  | :::| :|: ||| |  :|  ||
Db       4993 CDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSPERLDAVLRPKVDAAWYLH 5052

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   |:||||:|| |  ||  ||:|||| ||:|||||| |    ||| |:| | |
Db       5053 ELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHCRQAQGLPALSLASGLW 5110


RESULT 7
US-09-320-878-1
```

```
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4551
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-1

  Query Match             45.0%;  Score 428.5;  DB 3;  Length 4551;
  Best Local Similarity   48.9%;  Pred. No. 2.2e-39;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2146 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2204

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2205 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2264

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db       2265 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2322


RESULT 8
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             45.0%;  Score 428.5;  DB 4;  Length 4613;
  Best Local Similarity   48.9%;  Pred. No. 2.3e-39;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2208 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2266

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2267 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2326

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db       2327 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2384


RESULT 9
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             45.0%;  Score 428.5;  DB 4;  Length 11877;
  Best Local Similarity   48.9%;  Pred. No. 8.9e-39;
  Matches   87;  Conservative   22;  Mismatches   68;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||| :|    :  |   :|||| ||:|   | |  || :|   |  :||
Db       2544 GTVLITGGTGALGSHAARWM-AHHGAEHLLLVSRSGEQAPGATQLTAELTASGARVTIAA 2602

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  :: ||: || : ||| |:|||| |||  :  ||  |:    ||||  | :|
Db       2603 CDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRAHRAKAVGASVLD 2662

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|| ::| | ||| |||  |||||||  |      | |:||| |
Db       2663 ELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDALAARRRATGRSAVSVAWGPW 2720


RESULT 10
US-08-804-227C-9
; Sequence 9, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
```

```
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1996 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-9

  Query Match             44.4%;  Score 423;  DB 2;  Length 1996;
  Best Local Similarity   46.6%;  Pred. No. 2.9e-39;
  Matches   83;  Conservative   26;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||||||      ||       :||||  | ||  |     |  :| : |  : :
Db       1525 GTVLITGGTGALAAETARHLVDRHKVRHLLLVGRRGPDAPGVDRLVAELTESGAEVAVRA 1584

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: | |::||: :| :|||| |:||||:|||  |:  |  ::: ||| || :|  |
Db       1585 CDVTDRDALRRLLDALPDEHPLTCVVHTAGVLDDGVLSAQTAERIDTVLRPKADAAVHLD 1644

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:       ||||| :|| |: ||| ||||||::|||  |     || |: || |
Db       1645 ELTREIGRVPLVLYSSVSATLGSAGQAGYAAANAFMDALAARRCAAGHPALSLGWGWW 1702


RESULT 11
US-08-804-198-3
; Sequence 3, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1996 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-3

  Query Match             44.4%;  Score 423;  DB 2;  Length 1996;
  Best Local Similarity   46.6%;  Pred. No. 2.9e-39;
  Matches   83;  Conservative   26;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||||||      ||       :||||  | ||  |     |  :| : |  : :
Db       1525 GTVLITGGTGALAAETARHLVDRHKVRHLLLVGRRGPDAPGVDRLVAELTESGAEVAVRA 1584

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: | |::||: :| :|||| |:||||:|||  |:  |  ::: ||| || :|  |
Db       1585 CDVTDRDALRRLLDALPDEHPLTCVVHTAGVLDDGVLSAQTAERIDTVLRPKADAAVHLD 1644

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:       ||||| :|| |: ||| ||||||::|||  |     || |: || |
Db       1645 ELTREIGRVPLVLYSSVSATLGSAGQAGYAAANAFMDALAARRCAAGHPALSLGWGWW 1702


RESULT 12
US-08-804-227C-2
; Sequence 2, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4472 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-2

  Query Match             44.4%;  Score 423;  DB 2;  Length 4472;
  Best Local Similarity   47.2%;  Pred. No. 9.3e-39;
  Matches   84;  Conservative   22;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|| || |   |  ||       :|||| || |     |  |  :|
Db       4017 GTVLVTGATGTLGRLLARHLAAEHGVRHLLLLSRGGRAAEGADELAAELAGLEAEPCFAA 4076

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
```

```
                || :: : | ::|  :|   ||| ||| ||:||| ||  ||| ::  |:| || :|  ||
Db       4077 CDAADREALARVLAEVPADRPLTGVIHAAGVLDDGTLDALTPERIGTVMRPKADAALNLH 4136

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||: :||: | ::| ||    | ||||||||| :||||| ||   | ||| |:||| |
Db       4137 ELTRTSPLSVFAVFSGAAGILGRPGQANYAAANTFLDALAQHRRAHGLPAVSLAWGLW 4194


RESULT 13
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-10

  Query Match             44.2%;  Score 421;  DB 2;  Length 3724;
  Best Local Similarity   46.6%;  Pred. No. 1.2e-38;
  Matches   83;  Conservative   26;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||||| : :  || :    : ||: :| ||    |  |   |   |   |:
Db       3268 GTVLVTGGTGALGSRIARHLASRHGVRRLLIAARRGPDGEGAAELVADLAALGASATVVA 3327

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || |: | :: ||  ||  |||| |:|: |:|||  |  ||| ::  ||| |  :|  |
Db       3328 CDVSDADAVRGLLAGIPADHPLTAVVHSTGVLDDGVLPGLTPERMRRVLRPKVEAAVHLD 3387

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   |:||||:||:|   |:| | ||||||| |||||  | :  ||: |:||| |
Db       3388 ELTRDLDLSAFVLFSSSAGLLGSPAQGNYAAANATLDALAARRRSLGLPSVSLAWGLW 3445


RESULT 14
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-4

  Query Match             44.2%;  Score 421;  DB 2;  Length 3724;
  Best Local Similarity   46.6%;  Pred. No. 1.2e-38;
  Matches   83;  Conservative   26;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||||| : :  || :    : ||: :| ||    |  |   |   |   |:
Db       3268 GTVLVTGGTGALGSRIARHLASRHGVRRLLIAARRGPDGEGAAELVADLAALGASATVVA 3327

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || |: | :: ||  ||  |||| |:|: |:|||  |  ||| ::  ||| |  :|  |
Db       3328 CDVSDADAVRGLLAGIPADHPLTAVVHSTGVLDDGVLPGLTPERMRRVLRPKVEAAVHLD 3387

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||:   |:||||:||:|   |:| | ||||||| |||||  | :  ||: |:||| |
Db       3388 ELTRDLDLSAFVLFSSSAGLLGSPAQGNYAAANATLDALAARRRSLGLPSVSLAWGLW 3445


RESULT 15
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 6095
```

; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-144-085-2

Query Match 43.4%; Score 413.5; DB 4; Length 6095;
Best Local Similarity 48.0%; Pred. No. 1.8e-37;
Matches 86; Conservative 25; Mismatches 67; Indels 1; Gaps 1;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIH-LTIT 59
              || |:||||| |   :  ||      :||:|||| |   | |: |   |:: |   :|:
Db       1431 GTVLVTGGTGELGQAIAAHLVRAHGVRHLVLTSRRGLEAPGARELVQSLEELGAETVTVA 1490

Qy         60 TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 119
              || |  ::: ::|  |    ||: |:| ||:|||  ||  |  :|: ||  |   |  |
Db       1491 ACDVSKREEVARVLAGIDAARPLSAVLHLAGVLDDGVLTAQTAERLSRVLAPKVDGALHL 1550

Qy        120 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              |:||:   | ||||:|||| |||| ||:|||||| :|||| ||    | |||:||| |
Db       1551 HELTRELDLAAFVLFSSAAGTFGAAGQSNYAAANTFLDALAAHRRGGGLAATSLAWGFW 1609
```

Search completed: June 17, 2003, 13:09:27
Job time : 4.23524 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 8.71975 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_1715_1892
Perfect score: 953
Sequence: 1 GTTLITGGTGALATHLTHHL.....LAHHRHTHHLPATSIAWGTW 178

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	953	100.0	3972	21	AAB23749	S. avermitilis ave
2	953	100.0	3972	22	AAG65264	Streptomyces averm
3	953	100.0	3972	22	AAG65268	Streptomyces averm
4	948	99.5	6239	21	AAB23750	S. avermitilis ave
5	948	99.5	6239	22	AAG65265	Streptomyces averm
6	500	52.5	5435	22	AAE10145	Streptomyces nours
7	483	50.7	11096	22	AAE10129	Streptomyces nours
8	475	49.8	2066	22	AAE10146	Streptomyces nours
9	461.5	48.4	9477	22	AAE10144	Streptomyces nours
10	457	48.0	1864	18	AAW22602	Tylactone synthase
11	456.5	47.9	6797	22	AAB31558	Pimaricin biosynth
12	452	47.4	4881	21	AAB23751	S. avermitilis ave
13	452	47.4	4881	22	AAG65266	Streptomyces averm
14	441.5	46.3	7068	22	AAE10142	Streptomyces nours
15	439	46.1	4545	18	AAW22611	Hybrid srmG/tylG O
16	439	46.1	4550	18	AAW23716	Platenolide syntha
17	439	46.1	4550	18	AAW22606	Platenolide syntha
18	439	46.1	4572	19	AAW52845	A. mediterranei ri
19	436	45.8	5588	20	AAY39301	SpnE a polyketide
20	436	45.8	5588	22	AAB70969	S. spinosa protein
21	429.5	45.1	1557	21	AAY84730	Amino acid sequenc
22	429.5	45.1	1574	21	AAY84729	Amino acid sequenc
23	429.5	45.1	1578	21	AAY84732	Amino acid sequenc
24	429.5	45.1	1588	21	AAY84731	Amino aicd sequenc
25	429.5	45.1	1605	21	AAY84733	Amino acid sequenc
26	428.5	45.0	4551	21	AAB18637	Amino acid sequenc
27	428.5	45.0	4551	21	AAY67201	Narbonolide syntha
28	428.5	45.0	4613	21	AAY77192	S. venezuelae macr
29	428.5	45.0	4613	21	AAY77200	S. venezuelae pik
30	428.5	45.0	12199	21	AAY77180	S. venezuelae pik
31	425	44.6	3192	22	AAE10128	Streptomyces nours
32	423	44.4	1996	18	AAW22607	Platenolide syntha
33	423	44.4	1996	18	AAW23717	Platenolide syntha
34	423	44.4	4472	18	AAW22601	Tylactone synthase
35	421	44.2	3724	18	AAW23718	Platenolide syntha
36	421	44.2	3724	18	AAW22608	Platenolide syntha
37	415.5	43.6	3546	22	AAB82212	Polyketide synthas
38	415	43.5	4630	18	AAW19629	Streptomyces venez
39	413.5	43.4	6095	22	AAU10701	Amino acid sequenc
40	412	43.2	4630	21	AAY77177	S. venezuelae vep
41	411	43.1	3729	18	AAW22603	Tylactone synthase
42	411	43.1	3739	21	AAB18638	Amino acid sequenc
43	411	43.1	3739	21	AAY77193	S. venezuelae macr
44	411	43.1	3739	21	AAY77201	S. venezuelae pik
45	411	43.1	3739	21	AAY67202	Narbonolide syntha

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 953; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1.7e-93;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
Db 1715 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 1774

Qy 61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
Db 1775 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 1834

Qy 121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
Db 1835 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 1892

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 953; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1.7e-93;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
Db 1715 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 1774

Qy 61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
Db 1775 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 1834

Qy 121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
Db 1835 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 1892

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.

```
XX
SQ   Sequence   3972 AA;

  Query Match           100.0%;  Score 953;  DB 22;  Length 3972;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-93;
  Matches  178;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1715 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 1774

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1775 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 1834

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1835 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 1892


RESULT 4
AAB23750
ID   AAB23750 standard; Protein; 6239 AA.
XX
AC   AAB23750;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92301.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 223-254; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   6239 AA;

  Query Match           99.5%;  Score 948;  DB 21;  Length 6239;
  Best Local Similarity  99.4%;  Pred. No. 1.1e-92;
  Matches  177;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2306 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 2365

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2366 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 2425

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2426 QLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 2483


RESULT 5
AAG65265
ID   AAG65265 standard; Protein; 6239 AA.
XX
AC   AAG65265;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 5.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79277.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 180-201; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   6239 AA;

  Query Match           99.5%;  Score 948;  DB 22;  Length 6239;
  Best Local Similarity  99.4%;  Pred. No. 1.1e-92;
  Matches  177;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2306 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 2365

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2366 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 2425

Qy        121 QLTQHTPLTAPVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2426 QLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 2483


RESULT 6
AAE10145
ID   AAE10145 standard; Protein; 5435 AA.
XX
AC   AAE10145;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysJ.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic.
XX
OS   Streptomyces noursei.
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 260-263; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS encoding Streptomyces noursei nystatin gene, NysJ.
XX
SQ   Sequence   5435 AA;

  Query Match             52.5%;  Score 500;  DB 22;  Length 5435;
  Best Local Similarity   55.1%;  Pred. No. 4.9e-44;
  Matches   98;  Conservative   16;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||| ||  |  || |    :|||| | ||  | | |  :|   |  :|:
Db       1740 GTVLLTGGTGGLAGLLARHLVTEHGARHLLLAGRRGPDAPGAAALHAELTALGAEVTVAA 1799

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::   |  || |:| :|||| |:|||:||| ||| | | :|  ||| |  :|  ||
Db       1800 CDVADRTALAALLATVPAEHPLTAVVHTAGVLDDGTLTALNPDRLATVLRPKVDAAWHLH 1859
```

```
Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||:|  | |||||||| |   | ||||||||| | :||||| |||   |||||:||| |
Db       1860 DLTRHLDLAAFVLYSSTAGVMGGPGQANYAAGNTFLDALAAHRHALGLPATSLAWGAW 1917


RESULT 7
AAE10129
ID   AAE10129 standard; Protein; 11096 AA.
XX
AC   AAE10129;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysC.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          35..455
FT                   /label= KS3_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          546..858
FT                   /label= AT3_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          872..1073
FT                   /label= DH3_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1381..1628
FT                   /label= KR3_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1662..1735
FT                   /label= ACP3_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1757..2180
FT                   /label= KS4_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2291..2603
FT                   /label= AT4_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2617..2818
FT                   /label= DH4_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3124..3371
FT                   /label= KR4_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3407..3480
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3501..3924
FT                   /label= KS5_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4032..4346
FT                   /label= AT5_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4360..4561
FT                   /label= DH5_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4953..5239
FT                   /label= ER5_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          5248..5495
FT                   /label= KR5_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5528..5601
FT                   /label= ACP5_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          5623..6046
FT                   /label= KS6_domain
FT                   /note= "Ketosynthase (KS) domain"
```

```
FT   Domain          6166..6478
FT                   /label= AT6_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          6492..6704
FT                   /label= DH6_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          7038..7281
FT                   /label= KR6_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          7315..7388
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          7408..7831
FT                   /label= KS7_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          7939..8253
FT                   /label= AT7_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          8267..8470
FT                   /label= DH7_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          8812..9086
FT                   /label= KR7_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          9120..9193
FT                   /label= ACP7_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          9214..9637
FT                   /label= KS8_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          9758..10072
FT                   /label= AT8_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          10086..10289
FT                   /label= DH8_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          10657..10904
FT                   /label= KR8_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          10939..11012
FT                   /label= ACP8_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17184.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 170-176; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin
CC   gene, NysC.
XX
SQ   Sequence   11096 AA;

  Query Match             50.7%;  Score 483;  DB 22;  Length 11096;
  Best Local Similarity   54.5%;  Pred. No. 9.5e-42;
  Matches   97;  Conservative   19;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |    |  ||      :|||| || ||    |  || :|   |  : |
Db       1381 GTVLITGGTGGLGAVLARHLVATHGVRHLLLASRRGPAADGADDLTAELTGLGATVHIAA 1440

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::|  |  || |:|  |||| |:||||::||  | :||| :|: ||| || :|  ||
Db       1441 CDVADPAALADLLGTVPAGHPLTVVVHTAGVVDDGVLGSLTPQRLDTVLRPKADAAWHLH 1500

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |:|  | ||||:|| ||| |:|||||||| ||:|||||  |    |||||:||| |
Db       1501 EATRHLDLDAFVLFSSVAATLGSPGQANYAAGNAFLDALAARRAATGLPATSLAWGPW 1558


RESULT 8
AAE10146
ID   AAE10146 standard; Protein; 2066 AA.
XX
AC   AAE10146;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysK.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          34..457
FT                   /label= KS18_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          568..881
FT                   /label= AT18_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          898..1102
FT                   /label= DH18_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1416..1663
FT                   /label= KR18_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1695..1769
FT                   /label= ACP18_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1849..2066
FT                   /label= TE_domain
FT                   /note= "Thioesterase (TE) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
```

```
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 263-264; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin gene,
CC   NysK.
XX
SQ   Sequence   2066 AA;

  Query Match           49.8%;  Score 475;  DB 22;  Length 2066;
  Best Local Similarity   52.8%;  Pred. No. 6.5e-42;
  Matches   94;  Conservative   17;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || ||||||| |   |  || | :  :||||  | ||  | |  |  :|  :|  : :
Db       1416 GTVLITGGTGGLGAALARHLVTVRGARHLLLAGRRGPEAPGAGELVAELTAQGADVRVAA 1475

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              ||  :   |  || |:|  |||| |:||||:|||| : :||| ||  ||| || :|  ||
Db       1476 CDVGDRTALDALLATVPAAHPLTAVVHTAGVLDDALIGSLTPDQLATVLRPKADAAWHLH 1535

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
               |:   |  |||||| :   |:||| |||||||||||| ||    ||| |:||| |
Db       1536 DATRGLDLAGFVLYSSVSGVLGSPGQGNYAAANAYLDALARHRADQGLPALSLAWGPW 1593


RESULT 9
AAE10144
ID   AAE10144 standard; Protein; 9477 AA.
XX
AC   AAE10144;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (complete).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          41..464
FT                   /label= KS15_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          578..889
FT                   /label= AT15_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          903..1102
FT                   /label= DH15_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1446..1731
FT                   /label= ER15_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          1740..1988
FT                   /label= KR15_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          2023..2096
FT                   /label= ACP15_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          2117..2538
FT                   /label= KS16_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2635..2953
FT                   /label= AT16_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2967..3167
FT                   /label= DH16_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3257..3500
FT                   /label= KR16_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3539..3612
FT                   /label= ACP16_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3634..4057
FT                   /label= KS17_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4153..4472
FT                   /label= AT17_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4486..4725
FT                   /label= DH17_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4997..5245
FT                   /label= KR17_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5277..5350
FT                   /label= ACP17_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 255-260; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
```

```
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (complete).
XX
SQ   Sequence   9477 AA;

  Query Match             48.4%;  Score 461.5;  DB 22;  Length 9477;
  Best Local Similarity   52.2%;  Pred. No. 1.7e-39;
  Matches   93;  Conservative   20;  Mismatches   64;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:|||||||  |:   |      |||:| || ||  | |  |  :|   |  :|:
Db       9050 GTVLVTGGTGALGAHVARRL-AKDGAQHLVLLSRRGPDAPGAAELRAELDALGTDVTVAA 9108

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: |||  :|: :|   ||| |:||||:|||  |  ||| :   | |||  || ||
Db       9109 CDVADRDQLTAVLDALPADRPLTGVVHTAGVLDDGVLDRLTPERFQEVFRAKVTSALLLD 9168

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :||:   | ||||:|||:|  | ||||||||||| |||||  |    |||||::|| |
Db       9169 ELTRDRELAAFVLFSSASAAVGNPGQANYAAANAVLDALAEQRRVLGLPATSVSWGAW 9226


RESULT 10
AAW22602
ID   AAW22602 standard; Protein; 1864 AA.
XX
AC   AAW22602;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase ORF2 protein.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "encoded by GTG"
FT   Domain          21..449
FT                   /note= "ketosynthase domain, KS3"
FT   Domain          569..924
FT                   /note= "acyltransferase domain, AT3"
FT   Domain          950..1148
FT                   /note= "dehydratase domain, DH3"
FT   Domain          1412..1596
FT                   /note= "ketoreductase domain, KR3"
FT   Domain          1689..1772
FT                   /note= "acyl carrier protein domain, ACP3"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80413.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 9; Pages 80-86; 220pp; English.
XX
CC   AAW22601-W22605 represent proteins encoded by the tylactone synthase
CC   gene cluster of the invention. The gene cluster is also referred to as
CC   the tylG gene, and was isolated from Streptomyces fradiae. These
CC   sequences are multifunctional proteins which direct the synthesis of the
CC   polyketide tylactone, isolated from Streptomyces fradiae. Tylactone is
CC   the basic building block of the antibiotic tylosin. The DNA sequence can
CC   be modified so as to alter the type of carboxylic acids incorporated,
CC   the number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence   1864 AA;

  Query Match             48.0%;  Score 457;  DB 18;  Length 1864;
  Best Local Similarity   50.6%;  Pred. No. 5.1e-40;
  Matches   90;  Conservative   22;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||:|       || |    : ||||||:||  | |  |  :|   | |  :
Db       1414 GTVLITGGTGSLGAEAARHLVTRHGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAA 1473

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || ::   |: ||: :|  |||| |:||||:|||  ||  || :|  ||| || :   ||
Db       1474 CDVADRAALRALLDRVPAGHPLTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLH 1533

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              :|||   |:||:||||||   |: ||: |||||||||: |  | :  ||| |: || |
Db       1534 ELTQGHALSAFILYSSAAGVLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPW 1591


RESULT 11
AAB31558
ID   AAB31558 standard; Protein; 6797 AA.
XX
AC   AAB31558;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase polypeptide.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin.
XX
OS   Streptomyces natalensis.
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   N-PSDB; AAF24892.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds  -
XX
PS   Disclosure; Page 81-101; 116pp; English.
XX
CC   The present sequence represents a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
```

```
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence   6797 AA;

  Query Match            47.9%;  Score 456.5;  DB 22;  Length 6797;
  Best Local Similarity  52.3%;  Pred. No. 3.6e-39;
  Matches   92;  Conservative   18;  Mismatches   65;  Indels    1;  Gaps    1;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |:||||||| || | |      :||:|| | || | | |   |  :|: :|:
Db       1191 GTVLVTGGTGALGAHLAHWL-ADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAA 1249

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: | |  :|  |||  ||| |:| ||:|||  |  ||| :   ||| || :|  ||
Db       1250 CDAADRDALAAVLADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLH 1309

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 176
             :|||   |  |||:||     |  ||||||||||||||| ||    |||||::||
Db       1310 ELTQDLDLDHFVLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWG 1365


RESULT 12
AAB23751
ID   AAB23751 standard; Protein; 4881 AA.
XX
AC   AAB23751;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:5.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 254-278; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   4881 AA;

  Query Match            47.4%;  Score 452;  DB 21;  Length 4881;
  Best Local Similarity  50.3%;  Pred. No. 7e-39;
  Matches   89;  Conservative   21;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          2 TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTC 61
              | |||||||  |   :  || |    :|||| :| |     |  |  :|   |   |:  |
Db       4184 TVLITGGTGVLGRLVARHLATAHGVRHLLLATRRGTAADGAADLVAELAGLGAEATVAAC 4243

Qy         62 DTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQ 121
              |  :   :  ||: :| ||||  ||||||::||  |:|:||| ::  || |||  |  ||
Db       4244 DIGDRAAVAALLDQVPAQHPLKAVIHTAGVVDDGILTSLTPERMEAVLHAKAFGAAHLHD 4303

Qy        122 LTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              ||:   || | ::||||||:||:||| || ||||:||||  ||||  ||  |:||| |
Db       4304 LTRDAGLTTFTVFSSAAASFGSPGQGNYTAANAFLDALMQHRHTQALPGRSLAWGLW 4360


RESULT 13
AAG65266
ID   AAG65266 standard; Protein; 4881 AA.
XX
AC   AAG65266;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 6.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79278.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 201-217; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   4881 AA;

  Query Match            47.4%;  Score 452;  DB 22;  Length 4881;
  Best Local Similarity  50.3%;  Pred. No. 7e-39;
  Matches   89;  Conservative   21;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          2 TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTC 61
              | |||||||  |   :  || |    :|||| :| |     |  |  :|   |   |:  |
Db       4184 TVLITGGTGVLGRLVARHLATAHGVRHLLLATRRGTAADGAADLVAELAGLGAEATVAAC 4243
```

```
Qy         62 DTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQ 121
Db       4244 DIGDRAAVAALLDQVPAQHPLKAVIHTAGVVDDGILTSLTPERMEAVLHAKAFGAAHLHD 4303

Qy        122 LTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
Db       4304 LTRDAGLTTFTVFSSAAASFGSPGQGNYTAANAFLDALMQHRHTQALPGRSLAWGLW 4360
```

RESULT 14
AAE10142
ID AAE10142 standard; Protein; 7068 AA.
XX
AC AAE10142;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin gene, NysI (partial).
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; PKS type I.
XX
OS Streptomyces noursei.
XX

FH	Key	Location/Qualifiers
FT	Domain	34..448
FT		/label= KS9_domain
FT		/note= "Ketosynthase (KS) domain"
FT	Domain	572..890
FT		/label= AT9_domain
FT		/note= "Acyltransferase (AT) domain"
FT	Domain	904..1123
FT		/label= DH9_domain
FT		/note= "Dehydratase (DH) domain"
FT	Domain	1443..1686
FT		/label= KR9_domain
FT		/note= "Ketoreductase (KR) domain"
FT	Domain	1720..1793
FT		/label= ACP9_domain
FT		/note= "Acyl carrier protein (ACP) domain"
FT	Domain	1813..2236
FT		/label= KS10_domain
FT		/note= "Ketosynthase (KS) domain"
FT	Domain	2346..2664
FT		/label= AT10_domain
FT		/note= "Acyltransferase (AT) domain"
FT	Domain	2678..2890
FT		/label= DH10_domain
FT		/note= "Dehydratase (DH) domain"
FT	Domain	2983..3229
FT		/label= KR10_domain
FT		/note= "Ketoreductase (KR) domain"
FT	Domain	3266..3339
FT		/label= ACP10_domain
FT		/note= "Acyl carrier protein (ACP) domain"
FT	Domain	3358..3780
FT		/label= KS11_domain
FT		/note= "Ketosynthase (KS) domain"
FT	Domain	3898..4217
FT		/label= AT11_domain
FT		/note= "Acyltransferase (AT) domain"
FT	Domain	4231..4432
FT		/label= DH11_domain
FT		/note= "Dehydratase (DH) domain"
FT	Domain	4523..4770
FT		/label= KR11_domain
FT		/note= "Ketoreductase (KR) domain"
FT	Domain	4806..4879
FT		/label= ACP11_domain
FT		/note= "Acyl carrier protein (ACP) domain"
FT	Domain	4801..5325
FT		/label= KS12_domain
FT		/note= "Ketosynthase (KS) domain"
FT	Domain	5432..5754
FT		/label= AT12_domain
FT		/note= "Acyltransferase (AT) domain"
FT	Domain	5768..5977
FT		/label= DH12_domain
FT		/note= "Dehydratase (DH) domain"
FT	Domain	6068..6315
FT		/label= KR12_domain
FT		/note= "Ketoreductase (KR) domain"
FT	Domain	6348..6421
FT		/label= ACP12_domain
FT		/note= "Acyl carrier protein (ACP) domain"
FT	Domain	6454..6873
FT		/label= KS13_domain
FT		/note= "Ketosynthase (KS) domain"

XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR N-PSDB; AAD17185.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 15; Page 181-185; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a PKS type I encoding Streptomyces noursei
CC nystatin gene, NysI (partial).
XX
SQ Sequence 7068 AA;

Query Match 46.3%; Score 441.5; DB 22; Length 7068;
Best Local Similarity 50.6%; Pred. No. 1.6e-37;
Matches 90; Conservative 20; Mismatches 67; Indels 1; Gaps 1;

```
Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
Db       6068 GTVLITGGTGGIGGHVARRLARDGAT-HLLLTSRRGPAAPGADALRAELEELGARVTLAA 6126

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
Db       6127 CDAADRDALAALLAELPDDAPLCAVFHTAGVVEDHVVDALTPENFAAVLRAKTVAAHHLH 6186

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
Db       6187 ELTADLDLAAFVLFSSTAGVLGAAGQGNYAAANAHLDALAEHRRSHGLTALSVAWGPW 6244
```

```
RESULT 15
AAW22611
ID   AAW22611 standard; Protein; 4545 AA.
XX
AC   AAW22611;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Hybrid srmG/tylG ORF1 protein.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; hybrid protein.
XX
OS   Streptomyces ambofaciens.
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "encoded by GTG"
FT   Domain          15..418
FT                   /note= "ketosynthase domain, KS'(s), from tylG"
FT   Domain          525..882
FT                   /note= "acyltransferase domain, AT(s), from tylG"
FT   Domain          942..1025
FT                   /note= "acyl carrier protein domain, ACP(s), from tylG"
FT   Domain          1060..1483
FT                   /note= "ketosynthase domain, KS1, from srmG"
FT   Domain          1596..1953
FT                   /note= "acyltransferase domain, AT1, from srmG"
FT   Domain          2232..2416
FT                   /note= "ketoreductase domain, KR1, from srmG"
FT   Domain          2533..2616
FT                   /note= "acyl carrier protein domain, ACP1, from srmG"
FT   Domain          2641..3064
FT                   /note= "ketosynthase domain, KS2, from srmG"
FT   Domain          3184..3520
FT                   /note= "acyltransferase domain, AT2, from srmG"
FT   Domain          3546..3727
FT                   /note= "dehydratase domain, DH2, from srmG"
FT   Domain          4083..4268
FT                   /note= "ketoreductase domain, KR2, from srmG"
FT   Domain          4374..4457
FT                   /note= "acyl carrier protein domain, ACP2, from srmG"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80415.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 23; Pages 198-212; 220pp; English.
XX
CC   This sequence represents a hybrid protein of the invention. This
CC   sequence was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with
CC   a EcoRI-SfuI fragment from tylG ORF1. The position of the residues from
CC   each of the two genes is not given in the specification. The srmG gene
CC   (see AAT80414) was isolated from Streptomyces ambofaciens, and encodes
CC   the multi-functional proteins which direct the synthesis of the
CC   polyketide platenolide. Platenolide is the basic building block of the
CC   macrolide antibiotic spiramycin. The tylG gene (see AAT80413) is the
CC   tylactone synthase gene cluster of the invention. The tylG sequence was
CC   isolated from Streptomyces fradiae, and encodes multifunctional proteins
CC   which direct the synthesis of the polyketide tylactone. Tylactone is the
CC   basic building block of the antibiotic tylosin. The hybrid sequence can
CC   be used to transform S. ambofaciens lacking the srmG ORF1 sequence, or S.
CC   fradiae lacking the tylG ORF1 sequence, so that they can produce
CC   polyketides. The DNA sequence can be modified so as to alter the type of
CC   carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence   4545 AA;

  Query Match             46.1%;  Score 439;  DB 18;  Length 4545;
  Best Local Similarity   49.4%;  Pred. No. 1.6e-37;
  Matches   88;  Conservative   23;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITT 60
              || |||||||||   :  ||      :|:|| || ||  | |  |  :|   |:  |
Db       4080 GTVLITGGTGALGRRVASHLARRYGVRHMLLVSRRGPDAPEAGPLERELAGLGVTATFLA 4139

Qy         61 CDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH 120
              || :: : ::: :  :|  |||| |:||||:|||  || ||  :|: ||| || :   ||
Db       4140 CDLTDIEAVRKAVAAVPSDHPLTGVVHTAGVLDDGALTGLTRQRLDTVLRPKADAVRNLH 4199

Qy        121 QLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 178
              : |   || ||||:|:||   | ||||:|||||| |||||  |    ||| |:||| |
Db       4200 EATLDRPLRAFVLFSAAAGLLGRPGQASYAAANAVLDALAGARRAAGLPAVSLAWGLW 4257


Search completed: June 17, 2003, 13:02:30
Job time : 9.71975 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 14.1889 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_1050_1356
Perfect score: 1598
Sequence: 1 VFVFPGQGGQWAGMGVRLLA.....AEDVTAIGSLRRGDNDTRRF 307

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1598	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	1598	100.0	5532	2	Q9S0R4	Q9s0r4 streptomyce
3	1366	85.5	6239	2	Q9S0R7	Q9s0r7 streptomyce
4	1227.5	76.8	3626	2	Q9F7T9	Q9f7t9 streptomyce
5	945	59.1	3970	2	Q93HI8	Q93hi8 streptomyce
6	891.5	55.8	3816	2	Q9KIV3	Q9kiv3 streptomyce
7	867	54.3	4685	2	Q93HJ2	Q93hj2 streptomyce
8	867	54.3	9510	2	Q93NX9	Q93nx9 streptomyce
9	866	54.2	3613	2	Q93HJ1	Q93hj1 streptomyce
10	861.5	53.9	4150	2	Q9KIV4	Q9kiv4 streptomyce
11	858	53.7	3190	2	Q93NW7	Q93nw7 streptomyce
12	852	53.3	2595	2	Q9ALM6	Q9alm6 saccharopol
13	848.5	53.1	3170	2	Q9ALM4	Q9alm4 saccharopol
14	848.5	53.1	5060	2	O52545	O52545 amycolatops
15	848.5	53.1	5069	2	O52789	O52789 amycolatops
16	844.5	52.8	1041	2	Q93C90	Q93c90 actinomadur
17	841.5	52.7	9507	2	Q9EWA1	Q9ewa1 streptomyce
18	841	52.6	1907	2	Q93HI9	Q93hi9 streptomyce
19	837.5	52.4	8817	2	Q53840	Q53840 polyangium
20	835.5	52.3	3654	2	O30766	O30766 streptomyce
21	832	52.1	3546	2	Q9F830	Q9f830 micromonosp
22	832	52.1	4881	2	Q9S0R3	Q9s0r3 streptomyce
23	828.5	51.8	3192	2	Q9L4W4	Q9l4w4 streptomyce
24	827	51.8	3562	2	Q9F829	Q9f829 micromonosp
25	826	51.7	9477	2	Q9L4X3	Q9l4x3 streptomyce
26	816	51.1	3524	2	Q93H86	Q93h86 streptomyce
27	815.5	51.0	4735	2	O54666	O54666 amycolatops
28	813	50.9	4840	2	Q93HJ4	Q93hj4 streptomyce
29	812.5	50.8	1762	2	O52546	O52546 amycolatops
30	812.5	50.8	1763	2	O52790	O52790 amycolatops
31	809.5	50.7	3201	2	Q9F828	Q9f828 micromonosp
32	807	50.5	7576	2	Q9ZGA4	Q9zga4 streptomyce
33	806.5	50.5	3413	2	O54593	O54593 amycolatops
34	806.5	50.5	4290	2	Q9WXC0	Q9wxc0 micromonosp
35	806	50.4	6146	2	Q93HJ5	Q93hj5 streptomyce
36	805	50.4	1728	2	O54591	O54591 amycolatops
37	801.5	50.2	4498	2	Q93HE5	Q93he5 streptomyce
38	800	50.1	3939	2	Q93HJ3	Q93hj3 streptomyce
39	795	49.7	7525	2	Q9KIE0	Q9kie0 streptomyce
40	790.5	49.5	2126	16	P94996	P94996 mycobacteri
41	786.5	49.2	6315	2	Q9ADL6	Q9adl6 polyangium
42	785	49.1	3352	2	Q93H83	Q93h83 streptomyce
43	784	49.1	4151	16	O53490	O53490 mycobacteri
44	775.5	48.5	1937	2	O30482	O30482 streptomyce
45	774	48.4	5588	2	Q9ALM2	Q9alm2 saccharopol

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 1598; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2.2e-105;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 1109

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 1169

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 1229

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 1289

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db -      1290 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 1349

Qy        301 DNDTRRF 307
              |||||||
Db       1350 DNDTRRF 1356
```

RESULT 2
Q9S0R4
ID Q9S0R4 PRELIMINARY; PRT; 5532 AA.
AC Q9S0R4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 3.
GN AVEA3.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84478.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 5532 AA; 575193 MW; EC7B2069D44A4057 CRC64;

Query Match 100.0%; Score 1598; DB 2; Length 5532;
Best Local Similarity 100.0%; Pred. No. 3.3e-105;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 662

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 722

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 782

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 842

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 902

Qy        301 DNDTRRF 307
              |||||||
Db        903 DNDTRRF 909
```

RESULT 3
Q9S0R7
ID Q9S0R7 PRELIMINARY; PRT; 6239 AA.
AC Q9S0R7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 2.
GN AVEA2.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84475.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000169; SHprot_acsite.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 4.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 6239 AA; 666283 MW; A706280940B502FA CRC64;

Query Match 85.5%; Score 1366; DB 2; Length 6239;
Best Local Similarity 81.8%; Pred. No. 1.6e-88;
Matches 251; Conservative 25; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||||||||||: || ||||||  : |||:|| ||| ||: ||| ||  |  |::|
Db       4815 VFVFPGQGGQWAGMGLTLLTSSPVFAEHIDACEKALTPWVPWSLTDILHRDPDDPAWQQA 4874

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||||||:|||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       4875 DVVQPVLFSIMVSLAALWRSYGIEPDAVLGHSQGEIAAAHICGALSLKDAAKTVALRSRA 4934

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |||||||| |||:|||||:|:||| ||| |:|||||:||| || |||| :|||||||:|
Db       4935 LAAVRGRGAMASLPLPAQDVQQLISERWEGQLWVAALNGPHSTTVSGDTKAVDEVLAHCT 4994

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
```

```
                ||:||:|||||||||||||||| :||| |||||:|||| ||||||||||||||||||||
Db       4995 DTGLRAKRIPVDYASHCPHVQPLHDELLHLLGDITPQPSTVPFFSTVEGTWLDTTTLDAA 5054

Qy        241 YWYRNLHQPVRPSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||||||||| |:| | ||||| |:|:|||||||||||||||:| |::|| ||||||
Db       5055 YWYRNLHQPVRPSHAIQTLTDDGHRAPIEISPHPTLVPAIEDTTENTTENITATGSLRRG 5114

Qy        301 DNDTRRF 307
              |||| ||
Db       5115 DNDTHRF 5121


RESULT 4
Q9F7T9
ID   Q9F7T9      PRELIMINARY;      PRT;  3626 AA.
AC   Q9F7T9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Avermectin polyketide synthase (Fragment).
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC31271;
RA   Hong Y.-S., Lee J.J.;
RT   "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT   and polyketide synthase gene from Streptomyces avermitilis.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF275943; AAG09812.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT   NON_TER    3626    3626
SQ   SEQUENCE   3626 AA;  380557 MW;  6272F5F088C1A8D0 CRC64;

  Query Match             76.8%;  Score 1227.5;  DB 2;  Length 3626;
  Best Local Similarity   85.1%;  Pred. No. 6.5e-79;
  Matches  245;  Conservative    5;  Mismatches   17;  Indels   21;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||| ||||||||||||||    |||||||||||||||||||||||||||||||||:|
Db       1037 VFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWEQA 1096

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||||||||||||||||||||: |||||: ||||||: ||||||||||||
Db       1097 DVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDAAKTVA----- 1151

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
                          || |||||||||||  |||||||||||||||||||||||| |||||||
Db       1152 -------------LPPQEVEQLIGER-GGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCA 1196

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              |||||  |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1197 GTGVRP-RIPVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAA 1255

Qy        241 YWYRNLHQPVRPSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTA 288
              |||||||||||||  |||||| |||| :||||||||||||||||||||
Db       1256 YWYRNLHQPVRPSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTA 1303


RESULT 5
Q93HI8
ID   Q93HI8      PRELIMINARY;      PRT;  3970 AA.
AC   Q93HI8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69199.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   3970 AA;  412657 MW;  A93E583FAFA68C47 CRC64;

  Query Match             59.1%;  Score 945;  DB 2;  Length 3970;
  Best Local Similarity   59.1%;  Pred. No. 1.2e-58;
  Matches  182;  Conservative   41;  Mismatches   83;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| |||||   || |:|||  |:: ||||| |:||||:: :||:: |   |:|
Db       2372 VFVFPGQGAQWAGMAAELLVSAPVFRARIEECEEALRPYVDWSLLAVLRQEPGTPPWDRV 2431

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||||||||:|||||||||||||||| ||||| | | |: ||||:||
Db       2432 DVVQPALWAVMVSLAALWRSYGVEPDAVLGHSQGEIAAACVCGALDLADGARVVALRARA 2491

Qy        121 L-AAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              : : : ||||| || ||:    :||   | ||: ||:|||| || |:|| :|:||:|| |
Db       2492 IGSTLAGRGGMVSVALPSAAARELIA-GWPGRVSVASVNGPVSTVVAGDPDALDELLARC 2550

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              |  |||||||||||||   |  :  :|   |. ::|: | :|||||: |  |||  |||
Db       2551 ADDGVRARRIPVDYASHTSQVAEIEGDLAAELAGLTPRSSAIPFFSTLTGAPLDTAKLDA 2610

Qy        240 AYWYRNLHQPVRPSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              |||:||   | |  | :|  | |  || |:|||| |||:::|| :||     |:|||||
Db       2611 GYWYQNLRHTVEFEAATRAALDQGFTVFAEISPHPVLVPSVQDTIDDTGRTAVAVGSLRR 2670

Qy        300 GDNDTRRF 307
              |   |||
Db       2671 DDGGARRF 2678


RESULT 6
Q9KIV3
ID   Q9KIV3      PRELIMINARY;      PRT;  3816 AA.
AC   Q9KIV3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   8,8a-deoxyoleandolide synthase 2.
GN   OLEAII.
OS   Streptomyces antibioticus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20363406; PubMed=10908114;
RA   Shah S., Xue Q., Tang L., Carney J.R., Betlach M., McDaniel R.;
RT   "Cloning, Characterization, and Heterologous Expression of a
RT   Polyketide Synthase and P-450 Oxidase Involved in the Biosynthesis of
RT   the Antibiotic Oleandomycin.";
RL   J. Antibiot. 53:502-508(2000).
DR   EMBL; AF220951; AAF82409.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3816 AA;  400235 MW;  6741BFD35F9B5C72 CRC64;

  Query Match             55.8%;  Score 891.5;  DB 2;  Length 3816;
  Best Local Similarity   58.6%;  Pred. No. 8e-55;
  Matches  180;  Conservative   31;  Mismatches   93;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||   || : ||||  :  |   | |   ||:|::||   |:||  |
Db        567 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLR--GGEAVLGRV 624

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||   || ||:|| ||:||||||||| | | ||| | |: | |||||
Db        625 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 684

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || |||  |  :: : : ||| ||||||| || |||||:|:||:|| |
Db        685 IARIAGGGGMVSVSLPAGRVRTML-DTYGGRLSVAAVNGPSSTVVSGDAQALDELLAGCE 743

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db        744 REGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTTALDAG 803

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| || | :    |  |:|||||
Db        804 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDTVEADAVALGSLRRD 863

Qy        301 DNDTRRF 307
              :   ||
Db        864 EGGLGRF 870


RESULT 7
Q93HJ2
ID   Q93HJ2      PRELIMINARY;      PRT;  4685 AA.
AC   Q93HJ2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA6.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69195.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW   Phosphopantetheine.
SQ   SEQUENCE   4685 AA;  489355 MW;  2FA34215A39A0DB8 CRC64;

  Query Match             54.3%;  Score 867;  DB 2;  Length 4685;
  Best Local Similarity   58.0%;  Pred. No. 5.9e-53;
  Matches  178;  Conservative   37;  Mismatches   90;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || |||  |   |||||  | ||| ||:| ||||: |:: ||||| :| |
Db       2096 VFVFPGQGSQWVGMGRDLWDVSPVFAESMVACERALSPLVDWSLRDVVFRDAGDPLWSRV 2155

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR- 119
              |||||||::|||||||:|||:|:|| ||:||||||:||| | | |||:| |: ||:|||
Db       2156 DVVQPVLWAVMVSLAAVWRSFGVEPAAVVGHSQGEVAAACVAGGLSLEDGARVVAVRSRL 2215

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              |  : |:||| || || : ||: :  |: ||: ||||||| |  |||: ||:||:|| |
Db       2216 VLEKLSGKGGMVSVALPVEGVEERLA-RFEGRIGVAAVNGPTSVVVSGEPEALDELLAEC 2274

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              |:||||||| ||||||   |  | ::||  |  | |: | | |:||| |  |||: |||
Db       2275 EGSGVRARRIAVDYASHSAQVDALNDDLLAELAGIRPKSSTVAFYSTVTGGRLDTSGLDA 2334

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||  || : |||    |: |||  | :||| |||| |  |: :| |      | :|||||
Db       2335 RYWVTNLRERVRFEPVVRLLADQAHGLFVEASPHPVLNMAVAETGEVADRAVVTVGSLRR 2394

Qy        300 GDNDTRR 306
              |   |
Db       2395 DDGGADR 2401


RESULT 8
Q93NX9
ID   Q93NX9      PRELIMINARY;      PRT;  9510 AA.
AC   Q93NX9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphI.
GN   AMPHI.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73501.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_6.
KW   Phosphopantetheine.
SQ   SEQUENCE   9510 AA;  992817 MW;  1A8E003A15D478CF CRC64;

  Query Match             54.3%;  Score 867;  DB 2;  Length 9510;
  Best Local Similarity   55.5%;  Pred. No. 1.4e-52;
  Matches  171;  Conservative   48;  Mismatches   87;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| |||||| :|| ||||| |:  |  ||| : ||:::|:||    |     ||
Db       3851 VFVFPGQGSQWAGMGAQLLDESPVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERV 3910

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |:||||||:||: |:||||||:|||||||||| | |||||:|  |: | ||::|
Db       3911 DVVQPASFAVMVSLAAVWRAQGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQA 3970

Qy        121 LA-AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :  :: ||||| || ||  |||  : | : ||: ||| |||||: |:|: ||:||: |
Db       3971 IGRSLAGRGGMMSVALPVAEVEARL-EAFEGRVSVAAENGPRSSVVAGEPEALDELHAQL 4029

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
                :||||: |||||| |||: | :|:|||| :::|: | :|||||| | |||||  :||
Db       4030 TAEEIRARRVAVDYASHSPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDA 4089

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||||:|   | |:|||: |    || |:||| || |  :::|  :|      | |:|||
Db       4090 GYWYRSLRGRVLFADAVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRR 4149

Qy        300 GDNDTRRF 307
                :   ||
Db       4150 DNGGLDRF 4157


RESULT 9
Q93HJ1
ID   Q93HJ1      PRELIMINARY;      PRT;  3613 AA.
AC   Q93HJ1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA7.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69196.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3613 AA;  379236 MW;  6FAAE5CAB158E442 CRC64;

  Query Match             54.2%;  Score 866;  DB 2;  Length 3613;
  Best Local Similarity   56.8%;  Pred. No. 5e-53;
  Matches  175;  Conservative   46;  Mismatches   85;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || |||  |  :||||  ||: | :||:|:||||: ::: |||||  :| |
Db        573 VFVFPGQGSQWVGMGRDLWDASPVFRARMEECGQALSPYVDWSLRNVVFRDAGDPLWSRV 632

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR- 119
              |||||||::|||||||:|||:|:|| ||:||||||:||| | | |||:| |: ||:||| .
Db        633 DVVQPVLWAVMVSLAAVWRSFGVEPAAVVGHSQGEVAAACVAGGLSLEDGARVVAVRSRL 692

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
                |  : |:||| || || : ||: :  |: ||: ||||||| |  |||: :|:||:|| |
Db        693 VLEKLSGKGGMVSVALPVEGVEERLA-RFEGRIGVAAVNGPGSAVVSGEPDALDELLAEC 751

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
                :||||||| ||||||   |  | ::||  | || |: | | |:||| |  |||: |||
Db        752 EESGVRARRIAVDYASHSAQVDALNDDLLAELADIRPKSSAVAFYSTVTGERLDTSGLDA 811

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||  || : |||   |> ||:  ::|||| |||| |  ||::| |       ||:|||||
Db        812 RYWVTNLRERVRFEPVVRLLAEQEYQVFVESSPHPVLTVAIQETQEGVFGTGTAVGSLRR 871

Qy        300 GDNDTRRF 307
                :   ||
Db        872 DEGGADRF 879


RESULT 10
Q9KIV4
ID   Q9KIV4      PRELIMINARY;      PRT;  4150 AA.
AC   Q9KIV4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   8,8a-deoxyoleandolide synthase 1.
GN   OLEAI.
OS   Streptomyces antibioticus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20363406; PubMed=10908114;
RA   Shah S., Xue Q., Tang L., Carney J.R., Betlach M., McDaniel R.;
RT   "Cloning, Characterization, and Heterologous Expression of a
RT   Polyketide Synthase and P-450 Oxidase Involved in the Biosynthesis of
RT   the Antibiotic Oleandomycin.";
RL   J. Antibiot. 53:502-508(2000).
```

```
DR   EMBL; AF220951; AAF82408.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4150 AA;  435261 MW;  9383296C4C16647D CRC64;

  Query Match             53.9%;  Score 861.5;  DB 2;  Length 4150;
  Best Local Similarity   56.7%;  Pred. No. 1.2e-52;
  Matches  174;  Conservative   35;  Mismatches   95;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||| || ||   || : ||||  :  |    |      ||:|::||   |:||  |
Db       3076 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDRLTGWSLVEVLR--GGEAVLGRV 3133

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||| || ||:|| ||:||||||||||| | | ||| | |: | |||||
Db       3134 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 3193

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || | |: |  :: : : ||: ||||||| || |||||:|:||:|| |
Db       3194 IARIAGGGGMVSVGLSAERVRTML-DTYGGRVSVAAVNGPSSTVVSGDAQALDELLAGCE 3252

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | |::|| | |||||||   |||| |||
Db       3253 REGVRARRVPVDYASHSAQMDQLRDELLEALADVTPQDSSVPFFSTVTADWLDTTALDAG 3312

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| | :| :    |  |: ||||
Db       3313 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGITETLDTFDADAVALSSLRRD 3372

Qy        301 DNDTRRF 307
              :    ||
Db       3373 EGGLDRF 3379


RESULT 11
Q93NW7
ID   Q93NW7      PRELIMINARY;      PRT;  3190 AA.
AC   Q93NW7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphB.
GN   AMPHB.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73513.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3190 AA;  334060 MW;  EC005DABF792D89B CRC64;

  Query Match             53.7%;  Score 858;  DB 2;  Length 3190;
  Best Local Similarity   55.2%;  Pred. No. 1.6e-52;
  Matches  170;  Conservative   44;  Mismatches   92;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||| || ||| |||  ||||| |:  |  ||| : || ::|:||   |    ||
Db       573 VFVFPGQGSQWVGMGARLLDESPVFAERIAECAAALAEFTDWDLLDVLRGAEGAPTLERV 632

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |:|||||||:||: |:|||||:|||||||||| | |||||:| |: |||||:|
Db       633 DVVQPASFAVMVSLAAVWRAQGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVALRSQA 692

Qy        121 LA-AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :   : ||||| || ||  :||  : | : ||: ||| |||||  |||:  |:||:
Db       693 IGRTLAGRGGMMSVALPVADVEARL-EAFGGRVSVAAENGPRSCVVSGEPGALDELSELL 751

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              ||| ||:||||||| ||: | :|:|:|| :::|: | || |||| | |||||:||
Db       752 TAEGVRVRRVPVDYASHSHHVEDLHDEILQLLAEVAPKASEVPLFSTVTGDWLDTTVMDA 811

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||:|:|   | |:||:: | |  || |:||| || |  :::|  :|      | |:|||
Db       812 GYWFRSLRGRVLFADAIRGLLDTEHRAFIEVSSHPVLAMSVQDMIDDAGVVGVASGTLRR 871

Qy        300 GDNDTRRF 307
              :    ||
Db       872 DNGGLDRF 879


RESULT 12
Q9ALM6
ID   Q9ALM6      PRELIMINARY;      PRT;  2595 AA.
AC   Q9ALM6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase loading and extender module 1.
GN   SPNA.
OS   Saccharopolyspora spinosa.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23264.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
```

```
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2595 AA;  273663 MW;  52EF0BBA4D14FA62 CRC64;

  Query Match             53.3%;  Score 852;  DB 2;  Length 2595;
  Best Local Similarity   55.2%;  Pred. No. 3.3e-52;
  Matches  169;  Conservative   47;  Mismatches   88;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||| :|| :| ||| |  | :|||| :|||::|:||  || |  :
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| ||::||||| |||||:|: | ||:||||||||| | ||||::|||: ||:|||
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              | |: | | |||:  ||:|| |::   |  |: || |||| || |||| ||: |:|| ||
Db        647 LTALAGSGAMASLQHPAEEVRQIL-LPWRDRIGVAGVNGPSSTLVSGDREAMAELLAECA 705

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :| |||||:|||| ||:: :|:|||.||  : |:   :| :||  |  ||   :||
Db        706 DRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPMDAD 764

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| ||| |  ||:||   |:  |:|:|||| |   |::|       :| |:|:|||| 
Db        765 YWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAGREVVALGTLRRG 824

Qy        301 DNDTRR 306
              :   |:
Db        825 EGGMRQ 830


RESULT 13
Q9ALM4
ID   Q9ALM4      PRELIMINARY;      PRT;  3170 AA.
AC   Q9ALM4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase extender modules 3-4.
GN   SPNC.
OS   Saccharopolyspora spinosa.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23266.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3170 AA;  334873 MW;  205D519FDFA3249C CRC64;

  Query Match             53.1%;  Score 848.5;  DB 2;  Length 3170;
  Best Local Similarity   56.2%;  Pred. No. 7.6e-52;
  Matches  172;  Conservative   35;  Mismatches   98;  Indels    1;  Gaps    1;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||| ||||| | || :|| ||| |  |  ||  :||||:||:||         :|
Db        563 VFVFPGQGWQWAGMAVDLLDASPTFARHMDECATALRRYVDWSLVDVLRGAENSPPLDRV 622

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:|||||| :|||||: | ||:||||||||||    | | |:|||: |||||||
Db        623 DVLQPASFAVMVSLAEVWRSYGVRPAAVVGHSQGEIAAACAAGVLPLEDAARLVALRSRA 682

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |  : |||||||:  || ||  |       ||| |||:|||||  |||| |||||:|| ||
Db        683 LKGLSGRGGMASLACPADEVAALFAGS-GGRLEVAAINGPRSVVVSGDLEAVDELLAECA 741

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              :||||||||||||  ||: :|  :|    :  :   ||::|||  | |:|    ||
Db        742 EKDMRARRIPVDYASHSAHVEVVRSPVLAAAAGVRHRDGQVPWWSTVIGDWVDPARLDGE 801

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| |||||  ||| | : |  :|:|:| || |  |:|:|  ::    | |:|:|||
Db        802 YWYRNLRQPVRFEHAVQGLVERGFGLFIEMSAHPVLTTAVEETGAESETAVAAVGTLRRD 861

Qy        301 DNDTRR 306
                ||
Db        862 SGGLRR 867


RESULT 14
O52545
ID   O52545      PRELIMINARY;      PRT;  5060 AA.
AC   O52545;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyketide synthase.
GN   RIFB.
OS   Amycolatopsis mediterranei (Nocardia mediterranei).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=33910;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RX   MEDLINE=98174059; PubMed=9512878;
RA   August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,
RA   Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
RA   Floss H.G.;
RT   "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
RT   the molecular analysis of the rif biosynthetic gene cluster of
RT   Amycolatopsis mediterranei S699.";
RL   Chem. Biol. 5:69-79(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RX   MEDLINE=98165773; PubMed=9497318;
RA   Kim C.G., Yu T.W., Fryhle C.B., Handa S., Floss H.G.;
RT   "3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
RT   formation of the precursor of mC7N units in rifamycin and related
RT   antibiotics.";
RL   J. Biol. Chem. 273:6030-6040(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RX   MEDLINE=21201076; PubMed=11278540;
RA   Yu T.-W., Mueller R., Mueller M., Zhang X., Draeger G., Kim C.-G.,
RA   Leistner E., Floss H.G.;
RT   "Mutational analysis and reconstituted expression of the biosynthetic
RT   genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the
RT   starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei
RT   S699.";
```

```
RL   J. Biol. Chem. 276:12546-12555(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RA   August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
RA   Floss H.G.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF040570; AAC01711.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5060 AA;  526542 MW;  13A6EEBE1A48A6B0 CRC64;

  Query Match             53.1%;  Score 848.5;  DB 2;  Length 5060;
  Best Local Similarity   58.4%;  Pred. No. 1.4e-51;
  Matches  180;  Conservative   34;  Mismatches   87;  Indels    7;  Gaps    5;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| ||||||  || || ||| |:  || ||  |||||: |:|| :|   : :|
Db        561 VWVFPGQGTQWAGMGRELLDSSEVFAARIAECETALGRWVDWSLTDVLRGEAD--LLDRV 618

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |:||| |||:| | |:||:||:|||||||||| | |||||:|||| |||||:|
Db        619 DVVQPASFAVMVGLAAVWASLGVEPEAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQA 678

Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |  ::    : | ||||: ||||||| |  ::|||||:|| |
Db        679 IAASLAGRGGMASVALSEEDATARL-EPWAGRVEVAAVNGPTSVVIAGDAEALDEALDAL 737

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
                ||| ||: ||||||  ||:  |: | |:|| |   |   |||:||| | |: |   ||
Db        738 DDQGVRIRRVAVDYASHTRHVEAARDALAEMLGGIRAQAPEVPFYSTVTGGWVEDAGVLD 797

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
              |||||| : |||  ||  | : |||||||| || ||  | :  :||  ||  |:||
Db        798 GGYWYRNLRRQVRFGPAVAELIEQGHRVFVEVSAHPVLVQPINELVDDTEAVVT--GTLR 855

Qy        299 RGDNDTRR 306
              | |   ||
Db        856 REDGGLRR 863


RESULT 15
O52789
ID   O52789      PRELIMINARY;      PRT;  5069 AA.
AC   O52789;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Rifamycin polyketide synthase, type 1.
OS   Amycolatopsis mediterranei (Nocardia mediterranei).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=33910;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LBG A3136;
RA   Schupp T., Toupet C., Engel N., Goff S.;
RT   "Cloning and sequence analysis of the putative rifamycin polyketide
RT   synthase gene cluster from Amycolatopsis mediterranei.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ223012; CAA11036.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5069 AA;  527694 MW;  C6E545CB77C0AA4F CRC64;

  Query Match             53.1%;  Score 848.5;  DB 2;  Length 5069;
  Best Local Similarity   58.4%;  Pred. No. 1.4e-51;
  Matches  180;  Conservative   34;  Mismatches   87;  Indels    7;  Gaps    5;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| ||||||  || || ||| |:  || ||  |||||: |:|| :|   : :|
Db        570 VWVFPGQGTQWAGMGRELLDSSEVFAARIAECETALGRWVDWSLTDVLRGEAD--LLDRV 627

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |:||| |||:| | |:||:||:|||||||||| | |||||:|||| |||||:|
Db        628 DVVQPASFAVMVGLAAVWASLGVEPEAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQA 687

Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |  ::    : | ||||: ||||||| |  ::|||||:|| |
Db        688 IAASLAGRGGMASVALSEEDATARL-EPWAGRVEVAAVNGPTSVVIAGDAEALDEALDAL 746

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
                ||| ||: ||||||  ||:  |: | |:|| |   |   |||:||| | |: |   ||
Db        747 DDQGVRIRRVAVDYASHTRHVEAARDALAEMLGGIRAQAPEVPFYSTVTGGWVEDAGVLD 806

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
              |||||| : |||  ||  | : |||||||| || ||  | :  :||  ||  |:||
Db        807 GGYWYRNLRRQVRFGPAVAELIEQGHRVFVEVSAHPVLVQPINELVDDTEAVVT--GTLR 864

Qy        299 RGDNDTRR 306
              | |   ||
Db        865 REDGGLRR 872


Search completed: June 17, 2003, 13:07:53
Job time : 15.1889 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 3.34793 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_1050_1356
Perfect score: 1598
Sequence: 1 VFVFPGQGGQWAGMGVRLLA.....AEDVTAIGSLRRGDNDTRRF 307

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	884.5	55.4	3567	1	ERY2_SACER	Q03132 saccharopol
2	858	53.7	3519	1	OL56_STRAT	Q07017 streptomyce
3	840.5	52.6	3491	1	ERY1_SACER	Q03131 saccharopol
4	790	49.4	3172	1	ERY3_SACER	Q03133 saccharopol
5	570	35.7	2110	1	MCAS_MYCBO	Q02251 mycobacteri
6	493	30.9	1538	1	PPSB_MYCTU	Q10978 mycobacteri
7	472	29.5	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	448.5	28.1	1774	1	MSAS_PENPA	P22367 penicillium
9	322.5	20.2	1986	1	WA_EMENI	Q03149 emericella
10	279.5	17.5	2181	1	STCA_EMENI	Q12397 emericella
11	275.5	17.2	312	1	FABD_HAEIN	P43712 haemophilus
12	269.5	16.9	317	1	FABD_BACSU	P71019 bacillus su
13	256	16.0	308	1	FABD_SALTY	O85140 salmonella
14	254	15.9	308	1	FABD_ECOLI	P25715 escherichia
15	243	15.2	2505	1	FAS_RAT	P12785 rattus norv
16	238	14.9	2109	1	PKS1_ASPPA	Q12053 aspergillus
17	231.5	14.5	293	1	FABD_SYNY3	P73242 synechocyst
18	222.5	13.9	2511	1	FAS_CHICK	P12276 gallus gall
19	178.5	11.2	2504	1	FAS_HUMAN	P49327 homo sapien
20	125	7.8	302	1	FABD_MYCTU	Q10501 mycobacteri
21	106	6.6	761	1	CTPA_MYCTU	Q10876 mycobacteri
22	101.5	6.4	722	1	VATI_HALN1	Q9hnd8 halobacteri
23	101.5	6.4	2073	1	FAS1_SCHPO	Q9uug0 s fatty aci
24	101	6.3	1245	1	NARZ_ECOLI	P19319 escherichia
25	99.5	6.2	387	1	TRAB_RHISN	P55416 rhizobium s
26	97	6.1	2037	1	FAS1_CANAL	P34731 c fatty aci
27	96	6.0	2076	1	FAS1_YARLI	P34229 y fatty aci
28	95.5	6.0	600	1	GIDA_DEIRA	Q9ryc3 deinococcus
29	95.5	6.0	693	1	VG50_BPMD2	O64240 mycobacteri
30	94.5	5.9	2051	1	FAS1_YEAST	P07149 s fatty aci
31	93.5	5.9	326	1	YGGR_ECOLI	P52052 escherichia
32	93	5.8	406	1	SYT2_AERPE	Q9yfy3 aeropyrum p
33	92	5.8	533	1	MCPE_RHIME	Q52877 rhizobium m
34	92	5.8	752	1	CTPB_MYCTU	Q10877 mycobacteri
35	90.5	5.7	629	1	SYR_PYRHO	O59147 pyrococcus
36	90.5	5.7	677	1	YI36_MYCTU	Q50597 mycobacteri
37	90	5.6	421	1	TRAB_AGRT5	Q44351 agrobacteri
38	89.5	5.6	403	1	HMDH_HALVO	Q59468 halobacteri
39	89	5.6	463	1	YI02_MYCTU	O53951 mycobacteri
40	89	5.6	2205	1	POLN_RUBVT	P13889 rubella vir
41	88	5.5	780	1	CTPA_MYCLE	P46839 mycobacteri
42	88	5.5	904	1	DPO1_MYCTU	Q07700 mycobacteri
43	87.5	5.5	990	1	TNP7_ECOLI	P13694 escherichia
44	86.5	5.4	788	1	TRS1_HCMVA	P09695 human cytom
45	86	5.4	418	1	Y824_DEIRA	Q9rw45 deinococcus

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             55.4%;  Score 884.5;  DB 1;  Length 3567;
  Best Local Similarity   57.3%;  Pred. No. 4.9e-63;
  Matches  177;  Conservative   41;  Mismatches   88;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||:||||| ||||||  ||:|||||| :::||:|::||  || | |:||:  |    :|
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||||:|||||| |||||||:|| ||:|||||||||||| |||:|:|||| |  |||
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : :: | ||||:| |    | : :   |  || |||||||||  |||:  |:       ||
Db        680 MRSLSGEGGMAAVALGEAAVRERL-RPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCA 738

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :| | | |||||| | :: :|||||  |||:|:|: | | ||||   :| | |||
Db        739 AEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTELDAR 798

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTE--DTAEDVTAIGSLR 298
              |||||| : |||:|||  ||: |:  |:|||||| :|:|:|:  |  | |||   :|||
Db        799 YWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAEDAVVVGSLH 858

Qy        299 RGDNDTRRF 307
              |   |   |
Db        859 RDGGDLSAF 867
```

```
RESULT 2
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;
```

```
  Query Match           53.7%;  Score 858;  DB 1;  Length 3519;
  Best Local Similarity 57.0%;  Pred. No. 6.7e-61;
  Matches  175;  Conservative   34;  Mismatches   94;  Indels    4;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR-RDAGDAVWER 59
              |||||||| || ||   || : ||||  :   |   | |   ||:|::|: |||   | |
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDA--TVLGR 2278

Qy         60 ADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
               ||||| |::|||||  || ||:|| ||:||||||||| | | ||| | |: | ||||
Db       2279 VDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSR 2338

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              |:| : | ||| || |||  |  :: : : ||: ||||||| || |||| :|:||:|| |
Db       2339 AIARIAGGGGMVSVSLPAGRVRTML-DTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGC 2397

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGLDISPQPSGVPFFSTVEGTWLDTTTLDA 239
                ||||||:|||||||    :  ||:|||| | ||:|| | |||||||   ||||| |||
Db       2398 EREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTTALDA 2457

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||: || : ||| :||: |    |   ||| |||| ||| || | :   ::   :|||||
Db       2458 GYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLRR 2517

Qy        300 GDNDTRR 306
                :   |
Db       2518 DEGGLDR 2524


RESULT 3
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match           52.6%;  Score 840.5;  DB 1;  Length 3491;
  Best Local Similarity 55.1%;  Pred. No. 1.7e-59;
  Matches  173;  Conservative   46;  Mismatches   84;  Indels   11;  Gaps    4;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAG----DAV 56
              |||||||| ||||| | || :|||||  :: | :|| | :|: |:  || :|     ||
Db       1031 VFVFPGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAA 1090

Qy         57 --WERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTV 114
                 || |||||:|:||||||::||::|:|| ||:||||||||| | ||||| |||: |
Db       1091 LSTERVDVVQPVMFAVMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVV 1150

Qy        115 ALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDE 174
              ||||| :| : |  ||||:  || ||   ||:    |: :||||||||  |:||:: :|
Db       1151 ALRSRVIATMPGNKGMASIAAPAGEVRARIGD----RVEIAAVNGPRSVVVAGDSDELDR 1206

Qy        175 VLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLG-DISPQPSGVPFFSTVEGTWLD 233
```

```
            ::| |    :||:|: ||||||  ||: :|: |   || |  | |  ||||||| | |
Db       1207 LVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDFHPLPGFVPFFSTVTGRWTQ 1266

Qy        234 TTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA 293
              ||| |||||| : |||:|||:|||: |:| |:||| || |  |||:  : :  |::|
Db       1267 PDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGDGSGADLSA 1326

Qy        294 IGSLRRGDNDTRRF 307
              | ||||||     |
Db       1327 IHSLRRGDGSLADF 1340


RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
```

```
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             49.4%;  Score 790;  DB 1;  Length 3172;
  Best Local Similarity   56.0%;  Pred. No. 1.8e-55;
  Matches  169;  Conservative   39;  Mismatches   84;  Indels   10;  Gaps    5;

Qy          3 VFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADV 62
              |||||| || ||   ||  | |||  :: || |||| ||||: |:|   :|    :| ||
Db        557 VFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDV 613

Qy         63 VQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALA 122
              ||| ||:||||||||||:|:|| ||:|||||||||| |||:|:|||| ||:||| |
Db        614 VQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLR 673

Qy        123 AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRS-TAVSGDAEAVDEVLAYCAG 181
               : |:|||||  |  ::  : || |:|| | :|:|||||   |:|::  :||::| |
Db        674 RLGGQGGMASFGLGTEQAAERIG-RFAGALSIASVNGPRSVVVVAGESGPLDELIAECEA 732

Qy        182 TGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAY 241
              :|||||||||| | |: ||||||  |  ||| : |  :|| |  :|| |:| ||
Db        733 EAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMDTAY 792

Qy        242 WYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA----IGSL 297
              || || : ||| || : ||: |   |||||||| |   ||  | |:|    |    :|:|
Db        793 WYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIE-ATLDSALPADAGACVVGTL 851

Qy        298 RR 299
              ||
Db        852 RR 853


RESULT 5
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             35.7%;  Score 570;  DB 1;  Length 2110;
  Best Local Similarity   41.3%;  Pred. No. 6.4e-38;
  Matches  124;  Conservative   49;  Mismatches  123;  Indels    4;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|| ||| ||| || :|||| ||||  :   |  :|    :|| : :
Db        534 VWVFSGQGSQWAAMGTQLLASEPVFAATIAKLEPVIAAESGFSVTEAI---TAQQTVTGI 590

Qy         61 DVVQPVLFSVMVSLAA-LWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
              | ||| :|:| |:||| : ::||: | ||:||| || ||| | |||||:|||: :  ||:
Db        591 DKVQPAVFAVQVALAATMEQTYGVRPGAVVGHSMGESAAAVVAGALSLEDAARVICRRSK 650

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
               :  : | | | || |||::|   :  |    : |: |  |:|| : | :: | :::|
Db        651 LMTRIAGAGAMGSVELPAKQVNSELMARGIDDVVVSVVASPQSTVIGGTSDTVRDLIARW 710

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
                 | || : || ||| | | |: ::|   | ||:|   ||::|    :  |
Db        711 EQRDVMAREVAVDVASHSPQVDPILDDLAAALADIAPMTPKVPYYSATLFDPREQPVCDG 770

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              |||  ||   |:|: |||| :||:||| :||||| |  |:| |       | |:  :||
Db        771 AYWVDNLRNTVQFAAAVQAAMEDGYRVFADVSPHPLLTHAVEQTGRSLDMSVAALAGMRR 830


RESULT 6
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
```

```
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             30.9%;  Score 493;  DB 1;  Length 1538;
  Best Local Similarity   35.7%;  Pred. No. 7.2e-32;
  Matches  114;  Conservative   56;  Mismatches  129;  Indels   20;  Gaps    6;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||: |:| |||||| :|||  | ||  :   |          :|: |::   |
Db        561 VFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLRDVI---ATGKELVGI 617

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
```

```
              : :|  |  : ::|  ||||||::|| |:||| ||:||| | |||:  :  :  | |:|
Db        618 EQIQLGLIGMQLTLTELWRSYGVQPDLVIGHSMGEVAAAVVAGALTPAEGLRVTATRARL 677

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : |:|||| : | |   | || :   :: |   | || | ::|  | :||::|
Db        678 MAPLSGQGGMALLGLDAAATEALIAD--YPQVTVGIYNSPRQTVIAGPTEQIDELIARVR 735

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT---L 237
                  | |: :: | | | :  |:  :   | |::|:   :   ||   |: |  |
Db        736 AQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIIST---TYADLHTQPIF 792

Qy        238 DAAYWYRNLHQPVRFSDAVQAL---ADDGHRVFVEVSPHPTLVPAIEDTTED------TA 288
              || :|  |:  ||||  |: :    ||  :  |:|:| || |  || || ||      :|
Db        793 DAEHWATNMRNPVRFQQAIASAGSGADGAYHTFIEISAHPLLTQAIADTLEDAHRPTKSA 852

Qy        289 EDVTAIGSLRRGDNDTRRF 307
                :||:|:|  :||  |
Db        853 AKYLSIGTLQRDADDTVTF 871
```

```
RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             29.5%;  Score 472;  DB 1;  Length 1876;
  Best Local Similarity   34.9%;  Pred. No. 4.5e-30;
  Matches  110;  Conservative   54;  Mismatches  135;  Indels   16;  Gaps    6;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||: |:| |||||| :|||  | ||  :   |          :|: |:|   |
Db        632 VFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLHDVL---ANGEELVGI 688

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              : :|  |  : ::|  || ||| || |:||| ||:||| | |||:  :  :  | |||
Db        689 EQIQLGLIGMQLALTELWCSYGVRPDLVIGHSMGEVAAAVVAGALTPAEGLRVTATRSRL 748

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : |:|||| : | |   | || :    :: :   | || | ::|  | :||::|
Db        749 MAPLSGQGGMALLELDAPTTEALIAD--FPQVTLGIYNSPRQTVIAGPTEQIDELIARVR 806

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT---L 237
                | |: :: | | | :  |:  :   | |::|:   :   ||   |: |  |
Db        807 AQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIIST---TYADLHTQPVF 863

Qy        238 DAAYWYRNLHQPVRFSDAVQAL---ADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDV--T 292
              || :|  |:  |||| |: :    ||  :  |:|:| || |  || ||           |
Db        864 DAEHWATNMRNPVRFQQAIASAGSGADGAYHTFIEISAHPLLTQAIIDTLHSAQPGARYT 923

Qy        293 AIGSLRRGDNDTRRF 307
              ::|:|:|  :|   |
Db        924 SLGTLQRDTDDVVTF 938


RESULT 8
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             28.1%;  Score 448.5;  DB 1;  Length 1774;
  Best Local Similarity   38.2%;  Pred. No. 3.3e-28;
  Matches  116;  Conservative   46;  Mismatches  127;  Indels   15;  Gaps    7;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|| | | ||  || :|:  :|||   :|  :| :   :  | :::||   ||  :| :
Db        567 VWVFSGHGAQWPDMGKQLI-HNPVFFAAIQPLDELIQAEIGLSPIELLR--TGD--FESS 621

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              | || : : : : |:|| :| || | ||:||| |||||: | ||||  : |  |  |:
Db        622 DRVQILTYVMQIGLSALLQSNGITPQAVIGHSVGEIAASVVAGALSPAEGALIVTRRALL 681

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
                | |:||| | ||: | |:::| |    | ||  : | |  |:|| | | |
Db        682 YRQVMGKGGMILVNLPSAETEEILGSR--SDLVVAIDSSPSSCVVAGDKELVAETAEALK 739

Qy        181 GTGVRARRIPVDYASHCPH----VQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 236
              ||:   :  | | | |     | |||: | | |  :||    |  :||
Db        740 ARGVKTFTVKSDIAFHSPTLNGLVDPLRDVLAETLSPVSP---NVKLYSTALADPRGQDL 796

Qy        237 LDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTA-EDVTAIG 295
              |  ||  |:   || : ||:|  :||:|:|:||| || :  :| :|  |   ||   |
Db        797 RDVEYWAGNMVNRVRLTSAVKAAVEDGYRLFLEVSTHPVVSHSINETLMDAGMEDFAVIP 856
```

```
Qy        296 SLRR 299
              :| |
Db        857 TLLR 860


RESULT 9
WA_EMENI
ID   WA_EMENI       STANDARD;      PRT;  1986 AA.
AC   Q03149;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Conidial green pigment synthase (EC 2.3.1.-).
GN   WA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93101122; PubMed=1465094;
RA   Mayorga M.E., Timberlake W.E.;
RT   "The developmentally regulated Aspergillus nidulans wA gene encodes a
RT   polypeptide homologous to polyketide and fatty acid synthases.";
RL   Mol. Gen. Genet. 235:205-212(1992).
CC   -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC       INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC       BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC       ASEXUAL SPORES (CONIDIA).
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Conidial green pigment biosynthesis.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65866; CAA46695.1; -.
DR   PIR; S28353; S28353.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN      529    582       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   DOMAIN      991   1024       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   DOMAIN     1650   1719       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1772   1841       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    548    548       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE   1001   1001       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1682   1682       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1804   1804       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1986 AA;  216634 MW;  74EF0940FF40EE9A CRC64;

  Query Match           20.2%;  Score 322.5;  DB 1;  Length 1986;
  Best Local Similarity 29.6%;  Pred. No. 5.6e-18;
  Matches   95;  Conservative   54;  Mismatches  137;  Indels   35;  Gaps    9;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACE--------EALAPWVDWSVVDILRRDAG 53
              |:| ||| |:| || :|      |   ::  :         :: | || |:
Db        912 FLFTGQGAQYAAMGKQLYEDCSHFRSAIEHLDCISQGQDLPSILPLVDGSL--------- 962

Qy         54 DAVWERAD-VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAK 112
                : | :  |||     | ::|:: | | || |  ||||| |: || :  | ||  |
Db        963 -PLSELSPVVVQLGTTCVQMALSSFWASLGITPSFVLGHSLGDFAAMNAAGVLSTSDTIY 1021

Qy        113 TVALRSRALA--AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAE 170
                 |:: |      |   | ::  |  ||:||: |:      :| :| |  | :||
Db       1022 ACGRRAQLLTERCQPGTHAMLAIKAPLVEVKQLLNEKVHD---MACINSPSETVISGPKS 1078

Qy        171 AVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEG- 229
              ::||:   |:  |::  : | || |   |:|: |:| : |  |:     ||| | : |
Db       1079 SIDELSRACSEKGLKSTILTVPYAFHSAQVEPILEDLEKALQGITFNKPSVPFVSALLGE 1138

Qy        230 --TWLDTTTLDAAYWYRNLHQPVRFSDAVQALAD----DGHRVFVEVSPHPTLVPAIEDT 283
                 |   :  |:| |  |:  : | |: | :|: :        :::|| ||      :: |
Db       1139 VITEAGSNILNAEYLVRHCRETVNFLSAFEAVRNAKLGGDQTLWLEVGPHTVCSGMVKAT 1198

Qy        284 TEDTAEDVTAIGSLRRGDNDT 304
                    | : |||| | ||
Db       1199 ---LGPQTTTMASLRR-DEDT 1215


RESULT 10
STCA_EMENI
ID   STCA_EMENI     STANDARD;      PRT;  2181 AA.
AC   Q12397;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN   STCA OR PKSST.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
RX   MEDLINE=96202293; PubMed=8643646;
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 4;
RX   MEDLINE=95370159; PubMed=7642507;
RA   Yu J.-H., Leonard T.J.;
RT   "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT   novel type I polyketide synthase.";
RL   J. Bacteriol. 177:4792-4800(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34740; AAC49191.1; -.
DR   EMBL; L39121; AAA81586.1; -.
DR   HSSP; P25715; 1MLA.
```

```
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW   Multifunctional enzyme.
FT   DOMAIN      383    814       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      884   1209       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1706   1777       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1830   1901       ACYL CARRIER (ACP) 2.
FT   DOMAIN        ?   2181       THIOESTERASE.
FT   ACT_SITE    552    552       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    978    978       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1738   1738       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1862   1862       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2028   2028       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2181 AA;  238831 MW;  5A3E5712AA9AD942 CRC64;

  Query Match          17.5%;  Score 279.5;  DB 1;  Length 2181;
  Best Local Similarity   26.9%;  Pred. No. 1.9e-14;
  Matches   82;  Conservative   62;  Mismatches  148;  Indels   13;  Gaps    7;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDI-LRRDAGDAVWER 59
              |   ||| |: ||| :|  | | |    ::  ::     :   | :::   :  ||:: :
Db        885 VIACSGQGGSQYTGMGWQLYNSYPTFRSDLERFDQLARSYGFPSPLEVYTSKPVGDSMEDL 944

Qy         60 ADV-VQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRS 118
                | ||  | |: ::|  |  |:|::| ||:||| || || :: | ||  |    | :|:
Db        945 LPVIVQLALVSLEMALGNLLGSFGLKPSAVIGHSLGEYAALYISGVLSAADTLYLVGMRA 1004

Qy        119 RALAAVRGRGGMASVPLPAQEVE--QLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 176
              : |      ||  | : : |  |   ::: |       ||  |||  | :||  : |  :
Db       1005 KLLQERCQRGTHAMLAVRASPVTLCEVLAE---SNCEVACHNGPNDTVLSGPLKEVMNLQ 1061

Qy        177 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTT- 235
              : ||::   : : :| |   |||: ||  :   ::    :|  | :    :|
Db       1062 NSLSATGIKGTLLKLPFAFHSAQVQPILEEFKNVARGVTFHKPQIPVLSPLLVKVIDEKG 1121

Qy        236 TLDAAYWYRNLHQPVRFSDAVQALADD----GHRVFVEVSPHPTLVPAIEDTTE-DTAED 290
              |:|  |  |:  :||:    ::   |        : ::| |   :   |: | : ||:
Db       1122 TVDPVYLARHCREPVKMVSVLEHARDQHIITDRTIVIDVGPKALMAGMIKTTLDKDTSSA 1181

Qy        291 VTAIG 295
              :  :|
Db       1182 LPTLG 1186


RESULT 11
FABD_HAEIN
ID   FABD_HAEIN     STANDARD;      PRT;   312 AA.
AC   P43712;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD OR HI0156.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32701; AAC21825.1; -.
DR   HSSP; P25715; 1MLA.
DR   TIGR; HI0156; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; Complete proteome.
FT   ACT_SITE     93     93       BY SIMILARITY.
FT   ACT_SITE    202    202       BY SIMILARITY.
SQ   SEQUENCE   312 AA;  33424 MW;  6261622B834AE878 CRC64;

  Query Match          17.2%;  Score 275.5;  DB 1;  Length 312;
  Best Local Similarity   28.5%;  Pred. No. 3.9e-15;
  Matches   91;  Conservative   47;  Mismatches  146;  Indels   35;  Gaps   12;

Qy          3 VFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADV 62
              |||||| |  ||    |    |:     :   :||  :  | :|     :  :  |:
Db          7 VFPGQGSQTVGMLADLATEYPIVIETFKQASDALG-YDLWYLVQQGPAEELNKTWQ---- 61

Qy         63 VQPVLFSVMVSLAALWRSY--GIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
               || | :  |::  :|:     ::|: : ||| || :|    | |  :|| | | || :
Db         62 TQPALLAASVAIYRVWKEKFPQLKPEVMAGHSLGEYSALVCAGVLDFQDAIKLVELRGKL 121

Qy        121 L--AAVRGRGGM-ASVPLPAQEVEQLIGERWAGRLWVAAV--NGPRSTAVSGDAEAVDEV 175
             :  |    | | | | : |  : :      :   | : |:||  | |    ::|   ||:
Db        122 MQQAVPEGTGAMYAIIGLDNEAIINACKQAEEGEV-VSAVNFNSPGQVVIAGAKAAVERA 180

Qy        176 LAYCAGTGV-RARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTV------E 228
              | |   |  || : |   |||  ::|  |:|   | :|      :   : |      |
Db        181 AALCKEAGAKRALPLAVSVPSHCALMKPAAEQLAVTLENIQINTPTISVLNNVDVKAETE 240

Qy        229 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTA 288
              || : | :      | |: |||::: |: :| ||  |  || |   |       |:
Db        241 GTEIRTALV------RQLYSPVRWTETVEKMAQDGVLVLAEVGPGKVL----NGLTKRIV 290

Qy        289 EDVTAIGSLRRGDNDTRRF 307
              |: ||       ||   |
Db        291 GDLQAISV-----NDVASF 304


RESULT 12
FABD_BACSU
ID   FABD_BACSU     STANDARD;      PRT;   317 AA.
AC   P71019; O34463;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
```

```
GN   FABD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96326321; PubMed=8759840;
RA   Morbidoni H.R., de Mendoza D., Cronan J.E. Jr.;
RT   "Bacillus subtilis acyl carrier protein is encoded in a cluster of
RT   lipid biosynthesis genes.";
RL   J. Bacteriol. 178:4794-4800(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98195738; PubMed=9534248;
RA   Foulger D., Errington J.;
RT   "A 28 kbp segment from the spoVM region of the Bacillus subtilis 168
RT   genome.";
RL   Microbiology. 144:801-805(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U59433; AAC44306.1; -.
DR   EMBL; Y13937; CAA74249.1; -.
DR   EMBL; Z99112; CAB13463.1; -.
DR   HSSP; P25715; 1MLA.
DR   SubtiList; BG11836; fabD.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; Complete proteome.
FT   ACT_SITE     91     91       BY SIMILARITY.
FT   ACT_SITE    201    201       BY SIMILARITY.
FT   CONFLICT    124    131       VPAGEGAM -> GCRLAKEQW (IN REF. 1).
SQ   SEQUENCE   317 AA;  34035 MW;  191AE828B1C91D7F CRC64;

  Query Match             16.9%;  Score 269.5;  DB 1;  Length 317;
  Best Local Similarity   28.4%;  Pred. No. 1.2e-14;
  Matches   81;  Conservative   47;  Mismatches  138;  Indels   19;  Gaps    8;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              |:||||| |: ||| |    |    |     :| |   :   : :      ||| |
Db          6 FLFPGQGSQFIGMGKELYEQVPAAKRLFDEADETLETKLSSLIFE------GDA--EELT 57

Qy         62 V---VQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRS 118
              :    || | :  :::   ::  || ||   ||| || :|    |||| |||  ||  |
Db         58 LTYNAQPALLTTSIAVLEKFKESGITPDFTAGHSLGEYSALVAAGALSFKDAVYTVRKRG 117

Qy        119 RAL--AAVRGRGGMASV-PLPAQEVEQLIGE--RWAGRLWVAAVNGPRSTAVSGDAEAVD 173
              :  |   | | ||::  : |: ::|:  :       : :| :| |    :|| |: |:
Db        118 EFMNEAVPAGEGAMAAILGMDAEALKQVTDKVTEEGNLVQLANLNCPGQIVISGTAKGVE 177

Qy        174 EVLAYCAGTGVRARRIPVDYAS--HCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTW 231
                  | : | ||:: :   |   ::|  |:| |:|     : :.||  | |
Db        178 LASELAKENGAK-RAIPLEVSGPFHSELMKPAAEKLKEVLDACDIKDADVPVISNVSADV 236

Qy        232 LDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTL 276
              :             |: |||| :::  |  :|   |:|: |   |
Db        237 MTEKADIKEKLIEQLYSPVRFEESINKLIAEGVTTFIEIGPGKVL 281


RESULT 13
FABD_SALTY
ID   FABD_SALTY     STANDARD;      PRT;   308 AA.
AC   O85140;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD OR STM1194.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=98317265; PubMed=9642179;
RA   Zhang Y., Cronan J.E. Jr.;
RT   "Transcriptional analysis of essential genes of the Escherichia coli
RT   fatty acid biosynthesis gene cluster by functional replacement with
RT   the analogous Salmonella typhimurium gene cluster.";
RL   J. Bacteriol. 180:3295-3303(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
```

```
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF044668; AAC38649.1; -.
DR   EMBL; AE008752; AAL20123.1; -.
DR   HSSP; P25715; 1MLA.
DR   StyGene; SG10637; fabD.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE     91     91       BY SIMILARITY.
FT   ACT_SITE    200    200       BY SIMILARITY.
SQ   SEQUENCE   308 AA;  32274 MW;  B3565B45341A611A CRC64;

  Query Match             16.0%;  Score 256;  DB 1;  Length 308;
  Best Local Similarity   28.9%;  Pred. No. 1.4e-13;
  Matches   87;  Conservative   47;  Mismatches  143;  Indels   24;  Gaps   11;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| |  ||   : |: |:          ||  :  |::      :  :  |:
Db          5 FVFPGQGSQSVGMLAEMAANYPIVEETFAEASAALG-YDLWALTQQGPAEELNKTWQ--- 60

Qy         62 VVQPVLFSVMVSLAALWRSYGIE-PDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
               || | :  |:|  :|:  | : |  : ||| || :|    | ::  || : | :| :
Db         61 -TQPALLTASVALWRVWQQQGGKMPALMAGHSLGEYSALVCAGVINFADAVRLVEMRGKF 119

Qy        121 L--AAVRGRGGM-ASVPLPAQEVEQLIGERWAGRLWVAAV--NGPRSTAVSGDAEAVDEV 175
              :  |   | ||| | : |    : :   |   |:: |: |  | |    ::|  |||:
Db        120 MQEAVPEGTGGMSAIIGLDDASIAKACEESAEGQV-VSPVNFNSPGQVVIAGHKEAVERA 178

Qy        176 LAYCAGTGV-RARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDT 234
               | |    |  ||  :||   |||  ::|  ::|   |  |:      ||  : |:
Db        179 GAACKAAGAKRALPLPVSVPSHCALMKPAADKLAVELAKITFSAPTVPVVNNVD----VK 234

Qy        235 TTLDAA----YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPT---LVPAIEDTTEDT 287
                |||        | |: ||::: :|: :|  |      || |       |   | ||   :
Db        235 CETDAAAIRDALVRQLYNPVQWTKSVEFIAAQGVEHLYEVGPGKVLTGLTKRIVDTLTAS 294

Qy        288 A 288
              |
Db        295 A 295


RESULT 14
FABD_ECOLI
ID   FABD_ECOLI     STANDARD;      PRT;   308 AA.
AC   P25715;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN   FABD OR TFPA OR B1092.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12;
RX   MEDLINE=92183950; PubMed=1339356;
RA   Magnuson K., Oh W., Larson T.J., Cronan J.E. Jr.;
RT   "Cloning and nucleotide sequence of the fabD gene encoding malonyl
RT   coenzyme A-acyl carrier protein transacylase of Escherichia coli.";
RL   FEBS Lett. 299:262-266(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92234941; PubMed=1314802;
RA   Verwoert I.I., Verbree E.C., van der Linden K.H., Nijkamp H.J.,
RA   Stuitje A.R.;
RT   "Cloning, nucleotide sequence, and expression of the Escherichia coli
RT   fabD gene, encoding malonyl coenzyme A-acyl carrier protein
RT   transacylase.";
RL   J. Bacteriol. 174:2851-2857(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95214624; PubMed=7700236;
RA   Bouquin N., Tempete M., Holland I.B., Seror S.J.;
RT   "Resistance to trifluoroperazine, a calmodulin inhibitor, maps to the
RT   fabD locus in Escherichia coli.";
RL   Mol. Gen. Genet. 246:628-637(1995).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [6]
RP   SEQUENCE OF 288-308 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92210530; PubMed=1556094;
RA   Cronan J.E. Jr., Rawlings M.;
RT   "The gene encoding Escherichia coli acyl carrier protein lies within
RT   a cluster of fatty acid biosynthetic genes.";
RL   J. Biol. Chem. 267:5751-5754(1992).
RN   [7]
RP   SEQUENCE OF 1-11.
RC   STRAIN=K12 / W3110;
RA   Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA   Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA   Hochstrasser D.F.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
RN   [8]
RP   SEQUENCE OF 1-10.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=95286570; PubMed=7768883;
RA   Serre L., Verbree E.C., Dauter Z., Stuitje A.R., Derewenda Z.S.;
RT   "The Escherichia coli malonyl-CoA:acyl carrier protein transacylase
RT   at 1.5-A resolution. Crystal structure of a fatty acid synthase
RT   component.";
RL   J. Biol. Chem. 270:12961-12964(1995).
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
```

```
CC       malonyl-[acyl-carrier protein].
CC   -!- PATHWAY: Fatty acid biosynthesis.
CC   -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC       TRANSACYLASE ACTIVITY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M87040; AAA23742.1; -.
DR   EMBL; Z11565; CAA77658.1; -.
DR   EMBL; M84991; AAA23738.1; -.
DR   EMBL; AE000210; AAC74176.1; -.
DR   EMBL; D90745; BAA35900.1; -.
DR   PIR; S20443; S20443.
DR   PIR; B41856; B41856.
DR   PDB; 1MLA; 25-JAN-95.
DR   SWISS-2DPAGE; P25715; COLI.
DR   EcoGene; EG11317; fabD.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   TIGRFAMs; TIGR00128; fabD; 1.
KW   Fatty acid biosynthesis; Transferase; 3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   ACT_SITE     91     91
FT   ACT_SITE    200    200
SQ   SEQUENCE   308 AA;  32286 MW;  F5901043D92FED8E CRC64;

  Query Match             15.9%;  Score 254;  DB 1;  Length 308;
  Best Local Similarity   29.1%;  Pred. No. 2.1e-13;
  Matches   87;  Conservative   47;  Mismatches  145;  Indels   20;  Gaps   11;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| |  ||   : || |:          ||  :  |::      :  :  |:
Db          5 FVFPGQGSQTVGMLADMAASYPIVEETFAEASAALG-YDLWALTQQGPAEELNKTWQ--- 60

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVL-GHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
               || | :  |:|  :|:  | :  |:: ||| || :|    | :   || : | :| :
Db         61 -TQPALLTASVALYRVWQQQGGKAPAMMAGHSLGEYSALVCAGVIDFADAVRLVEMRGKF 119

Qy        121 L--AAVRGRGGMAS-VPLPAQEVEQLIGERWAGRLWVAAV--NGPRSTAVSGDAEAVDEV 175
              :  |   | | ||: : |    : :   |   |:: |: |  | |    ::|  |||:
Db        120 MQEAVPEGTGAMAAIIGLDDASIAKACEEAAEGQV-VSPVNFNSPGQVVIAGHKEAVERA 178

Qy        176 LAYCAGTGV-RARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDT 234
              | |   |  ||  :||   |||  ::|  ::|   |  |:      ||  : |:
Db        179 GAACKAAGAKRALPLPVSVPSHCALMKPAADKLAVELAKITFNAPTVPVVNNVDVKC--E 236

Qy        235 TTLDAA--YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPT---LVPAIEDTTEDTA 288
              |  ||      | |: ||::: :|: :|  |    || |      |   | ||   :|
Db        237 TNGDAIRDALVRQLYNPVQWTKSVEYMAAQGVEHLYEVGPGKVLTGLTKRIVDTLTASA 295


RESULT 15
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match           15.2%;  Score 243;  DB 1;  Length 2505;
  Best Local Similarity   27.5%;  Pred. No. 1.9e-11;
  Matches   87;  Conservative   56;  Mismatches  149;  Indels   24;  Gaps   13;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              |:  | | || |||: |:      |   :   :||| | :   | |:|     :  ::  |
Db        494 FICSGMGTQWRGMGLSLMRLDS-FRESILRSDEALKP-LGVKVSDLL-LSTDEHTFD--D 548

Qy         62 VVQPV--LFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
              :|     | :: ::|  |  | |::|| ::||| ||:|  :  | || ::|      | :
Db        549 IVHSFVSLTAIQIALIDLLTSMGLKPDGIIGHSLGEVACGYADGCLSQREAVLAAYWRGQ 608

Qy        120 ALA-AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAY 178
               :  |    | ||:| |  :| :|    |    :  |  |   :  :||   ||:| :
Db        609 CIKDANLPAGSMAAVGLSWEECKQ----RCPPGVVPACHNSEDTVTISGPQAAVNEFVEQ 664

Qy        179 CAGTGVRARRIPV-DYASHCPHVQPLREELLELLGDI--SPQPSGVPFFST--VEGTWLD 233
                  || |: :     | |   :: :   ||: |  :   |:|    : ||   | |
Db        665 LKQEGVFAKEVRTGGLAFHSYFMEGIAPTLLQALKKVIREPRPRSARWLSTSIPEAQWQS 724

Qy        234 T--TTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDV 291
              :  |  | |   ||  || | :|:  : :  | | :|::||  |   ::   : :
Db        725 SLARTSSAEYNVNNLVSPVLFQEALWHVPE--HAVVLEIAPHALLQAVLKRGVKPSC--- 779

Qy        292 TAIGSLRRGDNDTRRF 307

                  | |  ::|   |   |
Db        780 TIIPLMKRDHKDNLEF 795

Search completed: June 17, 2003, 13:02:34
Job time : 5.34793 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 6.85529 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_1050_1356
Perfect score: 1598
Sequence: 1 VFVFPGQGGQWAGMGVRLLA.....AEDVTAIGSLRRGDNDTRRF 307

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	879.5	55.0	3573	2	S23070	erythronolide synt
2	858	53.7	3519	2	S43048	polyketide synthas
3	848.5	53.1	5069	2	T17464	rifamycin polyketi
4	840.5	52.6	3491	2	T43231	probable 6-deoxyer
5	828.5	51.8	2124	2	T28658	polyketide synthas
6	815.5	51.0	4735	2	T17463	rifamycin polyketi
7	812.5	50.8	1763	2	T17465	rifamycin polyketi
8	807	50.5	7576	2	T17428	FK506 polyketide s
9	806.5	50.5	3413	2	T17467	rifamycin polyketi
10	805	50.4	1728	2	T17466	rifamycin polyketi
11	790.5	49.5	2126	2	H70621	probable polyketid
12	790	49.4	3172	2	S22012	erythronolide synt
13	790	49.4	3178	2	S13595	6-deoxyerythronoli
14	784	49.1	4151	2	G70944	probable polyketid
15	775.5	48.5	1937	2	T03224	probable polyketid
16	772.5	48.3	1346	2	T17412	polyketide synthas
17	772.5	48.3	3739	2	T17410	polyketide synthas
18	772	48.3	6260	2	T30228	polyketide synthas
19	772	48.3	10223	2	T30225	polyketide synthas
20	769	48.1	4613	2	T17409	polyketide synthas
21	766.5	48.0	8563	2	T30226	polyketide synthas
22	763	47.7	1562	2	T17411	polyketide synthas
23	749	46.9	2723	2	T03221	probable polyketid
24	747	46.7	2100	2	T03223	probable polyketid
25	736	46.1	6420	2	T30283	polyketide synthas
26	575.5	36.0	2108	2	H70819	probable polyketid
27	570	35.7	2110	2	B44110	mycocerosate synth
28	570	35.7	2111	2	A70668	mycocerosate synth
29	561.5	35.1	1582	2	E70876	probable polyketid
30	561.5	35.1	2118	2	S72705	mycocerosate synth
31	559.5	35.0	1570	2	AC2012	hypothetical prote
32	559	35.0	2116	2	C86926	probable mycoceros
33	542.5	33.9	1293	2	T30871	orsellinic acid sy
34	532.5	33.3	1827	2	B70984	probable polyketid
35	530.5	33.2	2126	2	E70522	probable polyketid
36	500	31.3	635	2	S73017	polyketide synthas
37	500	31.3	1446	2	S73013	polyketide synthas
38	500	31.3	1540	2	H87203	polyketide synthas
39	500	31.3	1871	2	A87204	polyketide synthas
40	493.5	30.9	1815	2	S73021	polyketide synthas
41	493.5	30.9	1822	2	F87203	polyketide synthas
42	493	30.9	1538	2	E70874	probable ppsB prot
43	483	30.2	2188	2	A70984	probable polyketid
44	472.5	29.6	497	2	S37580	probable acyltrans
45	472.5	29.6	497	2	S37581	probable acyltrans

ALIGNMENTS

RESULT 1
S23070
erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase II
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C;Accession: S23070; S22011; S23205
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S23070
A;Molecule type: DNA
A;Residues: 1-3573 <BEV1>
A;Cross-references: EMBL:X62569
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22011
A;Molecule type: DNA
A;Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,1
3479,'DH',3480-3572 <BEV2>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44448.1; PID:g581651
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of t
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23205
A;Molecule type: protein
A;Residues: 2-12,'XXX' <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
A;Start codon: GTG
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ort-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; mult
F;52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;1404-1475/Domain: acyl carrier protein homology <ACP1>
F;1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;3420-3493/Domain: acyl carrier protein homology <ACP2>
F;1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

```
  Query Match             55.0%;  Score 879.5;  DB 2;  Length 3573;
  Best Local Similarity   57.0%;  Pred. No. 1.1e-59;
  Matches  176;  Conservative   41;  Mismatches   89;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||:||||| |||||| ||:|||||| :::||:|::||  || | |:||:  |    :|
Db        561 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 620

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||:|||||| ||||||:|| ||:|||||||||| || |||:|:|||| |  |||
Db        621 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAEHVAGALTLEDAAKLVVGRSRL 680

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : :: | ||||:| |    | : :   |  || |||||||||  |||:  |:       ||
Db        681 MRSLSGEGGMAAVALGEAAVRERL-RPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCA 739

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                |:| | | |||||| | :: :|||||  |||:|:|: | | ||||   :| | |||
Db        740 AEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTELDAR 799

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTE--DTAEDVTAIGSLR 298
              |||||| : |||:|||  ||: |:  |:||||| :| |:|:  |  | |||   :|||
Db        800 YWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAEDAVVVGSLH 859

Qy        299 RGDNDTRRF 307
              |   |   |
Db        860 RDGGDLSAF 868


RESULT 2
S43048
polyketide synthase type I - Streptomyces antibioticus
N;Contains: acyl carrier protein; acyltransferase; ketoacylsynthase; ketoreductase; thic
C;Species: Streptomyces antibioticus
C;Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 26-May-2000
C;Accession: S43048; S41729
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, February 1993
A;Reference number: S43048
A;Accession: S43048
A;Molecule type: DNA
A;Residues: 1-3519 <SWA>
A;Cross-references: EMBL:L09654; NID:g153407; PIDN:AAA19695.1; PID:g153408
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A;Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyketi
A;Reference number: S41729; MUID:94150470; PMID:8107683
A;Accession: S41729
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1683-3238,3273-3303,'K',3305-3407,'T',3409-3462,'Y',3464-3516,'E',3518-3519
A;Cross-references: EMBL:L09654
C;Genetics:
A;Start codon: GTG
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;570-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1202-1381/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1489-1560/Domain: acyl carrier protein homology <ACP1>
F;1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3143-3214/Domain: acyl carrier protein homology <ACP2>
F;3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

  Query Match             53.7%;  Score 858;  DB 2;  Length 3519;
  Best Local Similarity   57.0%;  Pred. No. 5.1e-58;
  Matches  175;  Conservative   34;  Mismatches   94;  Indels    4;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR-RDAGDAVWER 59
              |||||||| || ||   || : |||| :  |   | |   ||:|::|: |||   |  |

Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDA--TVLGR 2278

Qy         60 ADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
              ||||| |::|||||  || ||:|| ||:|||||||||| | | ||| | |: | ||||
Db       2279 VDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSR 2338

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              |:| : | ||| || |||  |  :: .: : ||: ||||||| || |||| :|:||:|| |
Db       2339 AIARIAGGGGMVSVSLPAGRVRTML-DTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGC 2397

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db       2398 EREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTTALDA 2457

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||: || : ||| :||: |  |   ||| |||| ||| || | :   ::   :|||||
Db       2458 GYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLRR 2517

Qy        300 GDNDTRR 306
              :   |
Db       2518 DEGGLDR 2524


RESULT 3
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             53.1%;  Score 848.5;  DB 2;  Length 5069;
  Best Local Similarity   58.4%;  Pred. No. 4.3e-57;
  Matches  180;  Conservative   34;  Mismatches   87;  Indels    7;  Gaps    5;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| ||||||  || || ||| |:  || ||  |||||: |:|| :|   : :|
Db        570 VWVFPGQGTQWAGMGRELLDSSEVFAARIAECETALGRWVDWSLTDVLRGEAD--LLDRV 627

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |:||| |||:| | |:||:||:|||||||||| | ||||||:|||| |||||:|
Db        628 DVVQPASFAVMVGLAAVWASLGVEPEAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQA 687

Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |  ::    : | ||||: ||||||| |  ::|||||:|| |
Db        688 IAASLAGRGGMASVALSEEDATARL-EPWAGRVEVAAVNGPTSVVIAGDAEALDEALDAL 746

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
              ||| ||: ||||||  ||:  |: | |:|| |  |   |||:||| | |: |   ||
Db        747 DDQGVRIRRVAVDYASHTRHVEAARDALAEMLGGIRAQAPEVPFYSTVTGGWVEDAGVLD 806

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
              ||||||  : |||  ||  | : |||||||| || ||  | :  :||   ||  |:||
Db        807 GGYWYRNLRRQVRFGPAVAELIEQGHRVFVEVSAHPVLVQPINELVDDTEAVVT--GTLR 864

Qy        299 RGDNDTRR 306
              | |   ||
Db        865 REDGGLRR 872
```

```
RESULT 4
T43231
probable 6-deoxyerythronolide B synthase, module 1/2 - Saccharopolyspora erythraea
C;Species: Saccharopolyspora erythraea
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000
C;Accession: T43231
R;Donadio, S.; Staver, M.J.; McAlpine, J.B.; Swanson, S.J.; Katz, L.
Science 252, 675-679, 1991
A;Title: Modular organization of genes required for complex polyketide biosynthesis.
A;Reference number: Z22354; MUID:91220065; PMID:2024119
A;Accession: T43231
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3491 <DON>
A;Cross-references: EMBL:M63676; NID:g152691; PID:g152692; PIDN:AAA26493.1
C;Genetics:
A;Gene: eryA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;57-336/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;416-483/Domain: acyl carrier protein homology <ACP1>
F;526-926/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1031-1317/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;1886-1957/Domain: acyl carrier protein homology <ACP2>
F;2001-2400/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2508-2791/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;3332-3403/Domain: acyl carrier protein homology <ACP3>
F;1921/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             52.6%;  Score 840.5;  DB 2;  Length 3491;
  Best Local Similarity   55.1%;  Pred. No. 1.1e-56;
  Matches  173;  Conservative   46;  Mismatches   84;  Indels   11;  Gaps    4;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAG----DAV 56
              |||||||| ||||| | || :|||||  :: | :|| | :|: |:  || :|     ||
Db       1031 VFVFPGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAA 1090

Qy         57 --WERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTV 114
                || ||||||:|:||||||::||::|:|| ||:||||||||| | ||||| ||| | .
Db       1091 LSTERVDVVQPVMFAVMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVV 1150

Qy        115 ALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDE 174
              ||||| :| : |  ||||:  || ||   ||:    |: :||||||||  |:||:: :|
Db       1151 ALRSRVIATMPGNKGMASIAAPAGEVRARIGD----RVEIAAVNGPRSVVVAGDSDELDR 1206

Qy        175 VLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLG-DISPQPSGVPFFSTVEGTWLD 233
              ::| |    :||:|: ||||||  ||: :|: |   || |  | |  |||||||  | |
Db       1207 LVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDFHPLPGFVPFFSTVTGRWTQ 1266

Qy        234 TTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA 293
                 ||| |||||| : |||:|||:|||: |:| |:||| || |  |||:  : :  |::|
Db       1267 PDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGDGSGADLSA 1326

Qy        294 IGSLRRGDNDTRRF 307
              | ||||||    |
Db       1327 IHSLRRGDGSLADF 1340


RESULT 5
T28658
polyketide synthase - Sorangium cellulosum (fragment)
C;Species: Sorangium cellulosum
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C;Accession: T28658
R;Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J.J.; Ligon, J.M.
J. Bacteriol. 177, 3673-3679, 1995
A;Title: A Sorangium cellulosum (myxobacterium) gene cluster for the biosynthesis of the
inomycetes.
A;Reference number: Z20499; MUID:95325306; PMID:7601830
A;Accession: T28658
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2124 <SCH>
A;Cross-references: EMBL:U24241; NID:g868150; PID:g868151; PIDN:AAA79984.1
A;Experimental source: strain So ce26
C;Function:
A;Description: involved in soraphen A biosynthesis
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;335-406/Domain: acyl carrier protein homology <ACP>
F;370/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             51.8%;  Score 828.5;  DB 2;  Length 2124;
  Best Local Similarity   54.2%;  Pred. No. 5.4e-56;
  Matches  166;  Conservative   45;  Mismatches   94;  Indels    1;  Gaps    1;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || || : || |||||  :::||| |||| |:||:: :|||| |    :|
Db        969 VFVFPGQGSQWEGMALSLLDSSPVFRAQLEACERALAPHVEWSLLAVLRRDEGAPSLDRV 1028

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| ||:||||||||||| |:|| ||:|||||||||| | |||||:|||:  ||||:|
Db       1029 DVVQPALFAVMVSLAALWRSLGVEPAAVVGHSQGEIAAAFVAGALSLEDAARIAALRSKA 1088

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |  | | | ||:| | | ::::  :    |  || :|||| ||:| |||:  |:| ::
Db       1089 LTTVAGNGAMAAVELGASDLQTYLAP-WGDRLSIAAVNSPRATLVSGEPAAIDALIDSLT 1147

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                | ||:: ||||||   :  :::||   | :|:|:   :| :||| || || : || |
Db       1148 AAQVFARKVRVDYASHSAQMDAVQDELAAGLANIAPRTCELPLYLSTVTGTRLDGSELDGA 1207

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| | | || | : | ||||| |||||||| |  |: :| | :  |  :||:||
Db       1208 YWYRNLRQTVLFSSATERLLDDGHRFFVEVSPHPVLTLALRETCERSPLDPVVVGSIRRD 1267

Qy        301 DNDTRR 306
              :   |
Db       1268 EGHLAR 1273


RESULT 6
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

  Query Match             51.0%;  Score 815.5;  DB 2;  Length 4735;
  Best Local Similarity   55.8%;  Pred. No. 1.5e-54;
  Matches  172;  Conservative   32;  Mismatches   89;  Indels   15;  Gaps    5;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| ||||||  ||| ||||| |:  |  |||||:|||:||:||   |:   |
Db       1144 VWVFPGQGTQWAGMGRELLAESPVFAERIAECAAALAPWIDWSLVDVLR---GEGDLGRV 1200

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:||| |||:| | |: ||||:|||||||||| | |||||:|||| |||||:|
Db       1201 DVLQPACFAVMVGLAAVWESVGVRPDAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQA 1260
```

```
Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |   :|      |    : ||||||| |  ::||| |:|  |
Db       1261 IAAELSGRGGMASVALGEDDVV----SRLVDGVEVAAVNGPSSVVIAGDAHALDATLEIL 1316

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
              :| |:| ||: ||||||  ||: :|: | | |  || |   |||:|||   |: |   ||
Db       1317 SGEGIRVRRVAVDYASHTRHVEDIRDTLAETLAGISAQAPAVPFYSTVTSEWVRDAGVLD 1376

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
               ||||||   |:|| |  || : || |||||| ||  |  : : | |      |||:||
Db       1377 GGYWYRNLRNQVRFGAAATALLEQGHTVFVEVSAHPVTVQPLSELTGD------AIGTLR 1430

Qy        299 RGDNDTRR 306
              | |   ||
Db       1431 REDGGLRR 1438
```

RESULT 7
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>

 Query Match 50.8%; Score 812.5; DB 2; Length 1763;
 Best Local Similarity 55.2%; Pred. No. 7.6e-55;
 Matches 170; Conservative 34; Mismatches 93; Indels 11; Gaps 5;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| || |||  || :||||| |:: |  ||  | |||::|:||   ||   :
Db        555 VWVFPGQGTQWVGMGRELLDASPVFAERIKECAAALDQWTDWSLLDVLR---GDGDLDSV 611

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              :|:||  |:||| |||:| | |: ||||:|||||||||| | |||:| |||| |||||:|
Db        612 EVLQPACFAVMVGLAAVWESAGVRPDAVVGHSQGEIAAACVSGALTLDDAAKVVALRSQA 671

Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |    :|  | ||: ||||||| |  ::|||:|:|| |
Db        672 IAARLSGRGGMASVALSEDEANARLG-LWDGRIEVAAVNGPASVVIAGDAQALDEALEVL 730

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
              || ||| |:: |||||| ||: :|: | | |  |: |   ||| ||| | |: |   ||
Db        731 AGDGVRVRQVAVDYASHTRHVEDIRDTLAETLAGITAQAPDVPFRSTVTGGWVRDADVLD 790

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
               ||||||   |||  ||  | : || |||||| || ||  | : |      |    |:||
Db        791 GGYWYRNLRNQVRFGPAVAELLEQGHGVFVEVSAHPVLVQPISELT-----DAVVTGTLR 845

Qy        299 RGDNDTRR 306
              | |   ||
Db        846 RDDGGLRR 853
```

RESULT 8
T17428
FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: T17428
R;Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressar
A;Reference number: Z18779; MUID:98451508; PMID:9780228
A;Accession: T17428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7576 <MOT>
A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A;Experimental source: strain MA6548
C;Genetics:
A;Gene: fkbB
C;Function:
A;Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein
F;54-500/Domain: acetate-CoA ligase homology <ACL>
F;1095-1166/Domain: acyl carrier protein homology <ACP1>
F;1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;2680-2751/Domain: acyl carrier protein homology <ACP2>
F;2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;4320-4391/Domain: acyl carrier protein homology <ACP3>
F;4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5903-5974/Domain: acyl carrier protein homology <ACP4>
F;6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;7450-7521/Domain: acyl carrier protein homology <ACP5>

 Query Match 50.5%; Score 807; DB 2; Length 7576;
 Best Local Similarity 55.7%; Pred. No. 1.2e-53;
 Matches 172; Conservative 34; Mismatches 89; Indels 14; Gaps 7;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || |||  |||: ||||||:  | |||||:  | ::|:: |  |    :|
Db       3295 VFVFPGQGSQWDGMGAELLATEPVFARRLGECAEALAPYTGWDLLDVIARRPGAPELDRV 3354

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |::||:|| |||::|: | ||:|||||:||| | | |:| |||| ||||||
Db       3355 DVVQPASFAMMVALAELWRAHGVAPAAVVGHSQGEVAAACVAGVLTLDDAAKVVALRSRL 3414

Qy        121 LAAVR-GRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|  | | ||| ||| ||           ||||| ||||||| |  |:| |:||:|:||
Db       3415 VATERAGHGGMVSVP-PA----DFDAAAWAGRLEVAAVNGPASIVVAGAADAVEELLA-- 3467

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
                  |||| ||||||  ||: :|  ||: | |::|    :||||||:  |||    ||
Db       3468 --ATPHARRIAVDYASHTAHVESIRGALLDALADLTPGAPEIPFFSTVDEAWLDRPA-DA 3524

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAI-GSLR 298
              |||| |:  ||||  |   ||: ||||||| |||| |  |: ||        ||: |:||
Db       3525 AYWYDNVRCPVRFGAAAARLAELGHRVFVEASPHPVLTTALADTL--AGHPNTAVTGTLR 3582

Qy        299 RGDNDTRRF 307
              |||   |||
Db       3583 RGDGGARRF 3591
```

RESULT 9
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1

```
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

  Query Match            50.5%;  Score 806.5;  DB 2;  Length 3413;
  Best Local Similarity  55.8%;  Pred. No. 4.9e-54;
  Matches  172;  Conservative   35;  Mismatches   90;  Indels   11;  Gaps    5;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| || |||  ||  ||||| |:  |  || ||: ||: |:||   ||   :|
Db       2228 VWVFPGQGTQWVGMGRELLEESPVFAERIAECAAALEPWIGWSLFDVLR---GDGDLDRV 2284

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:||| |||:| | |: |||||||||||||| | |||||:|||| |||||:|
Db       2285 DVLQPACFAVMVGLAAVWSSAGVVPDAVLGHSQGEIAAACVSGALSLEDAAKVVALRSQA 2344

Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |   :|      | |  : ||||||| |  ::|||:|:|| |
Db       2345 IAAKLSGRGGMASVALGEADVV----SRLADGVEVAAVNGPASVVIAGDAQALDETLEAL 2400

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
              :| |:|||: |||||  ||: : : | | |  :  ||| ||: | |: |   :|
Db       2401 SGAGIRARRVAVDYASHTRHVEDIEDTLAEALAGIDARAPLVPFLSTLTGEWIRDEGVVD 2460

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
              ||||||   |||  ||:||   || ||||:| || ||  | : |::||  ||  ||||
Db       2461 GGYWYRNLRGRVRFGPAVEALLAQGHGVFVELSAHPVLVQPITELTDETAAVVT--GSLR 2518

Qy        299 RGDNDTRR 306
              | |   ||
Db       2519 RDDGGLRR 2526


RESULT 10
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17466
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17466
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1728 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1573-1644/Domain: acyl carrier protein homology <ACP>

  Query Match            50.4%;  Score 805;  DB 2;  Length 1728;
  Best Local Similarity  54.5%;  Pred. No. 2.8e-54;
  Matches  168;  Conservative   36;  Mismatches   96;  Indels    8;  Gaps    5;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |:|||||| || |||  || :||||| |:  |  || ||:|||::|:||   |:   :|
Db        563 VWVFPGQGSQWVGMGRELLEASPVFAARIAECAAALEPWIDWSLLDVLR---GEGDLDRV 619

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |:||| |||:| | |: |||||||||||||| | |||||:|||| |||||:|
Db        620 DVVQPASFAVMVGLAAVWSSVGVVPDAVLGHSQGEIAAACVSGALSLQDAAKVVALRSQA 679

Qy        121 LAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              :|| : |||||||| |  ::    :   || |: ||||| | |  ::|||||:|: |
Db        680 IAAKLAGRGGMASVALSEEDAVARL-RHWADRVEVAAVNSPSSVVIAGDAEALDQALEAL 738

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL-DTTTLD 238
               |  :| ||: ||||||  ||: ::| | | |  |      :|||||: | |: :   :|
Db        739 TGQDIRVRRVAVDYASHTRHVEDIQEPLAEALAGIEAHAPTLPFFSTLTGDWIREAGVVD 798

Qy        239 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 298
              ||||||   | |  ||  |.  |||||||| || || ||    :||  |   ||||  .
Db        799 GGYWYRNLRNQVGFGPAVAELLGLGHRVFVEVSAHPVLVQAISAIADDT--DAVVTGSLR 856

Qy        299 RGDNDTRR 306
              | :   ||
Db        857 REEGGLRR 864


RESULT 11
H70621
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70621
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2126 <COL>
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06632.1; PID:g1839013
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks7
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I h
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;563-845/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1694-1875/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1976-2051/Domain: acyl carrier protein homology <ACP2>

  Query Match            49.5%;  Score 790.5;  DB 2;  Length 2126;
  Best Local Similarity  52.3%;  Pred. No. 4.9e-53;
  Matches  162;  Conservative   42;  Mismatches  101;  Indels    5;  Gaps    2;

Qy          2 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| || ||| :| |::|||| :  || ||   |:||::|:||   |    :| |
Db        564 FVFPGQGAQWLGMGAQLCATAPVFAEHIHRCERALREHVEWSLLDVLRGAPGAPGLDRVD 623

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
              |||| |::|||||| |||| |: ||||:||||||||||:| |||||:|||  |||||| |
Db        624 VVQPALWAVMVSLAELWRSVGVVPDAVIGHSQGEIAAAYVAGALSLRDAAAVVALRSRLL 683

Qy        122 AAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAG 181
              : | ||| |:     : |:|   :|  || :|||||  |  ::|: :|| |::  |
Db        684 VRLGGAGGMVSLACGQPQAEKL-ASQWGDRLNIAAVNGVSSVVLAGETDAVTELMQRCEA 742

Qy        182 TGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAY 241
              |:||||| ||||||   |  :||||:  |  | |: | | ||||| |  :|| ::| |
Db        743 EGIRARRIDVDYASHSAQVDAIREELIAALRGIEPRTSTVAFFSTVTGELMDTAGVNAEY 802

Qy        242 WYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVT----AIGSL 297
              |||:: |||:|  ||:   | |:||||| |||| |:  ||:|  |      |     | :|
Db        803 WYRSIRQPVQFERAVRNAFDGGYRVFVESSPHPVLIAGIEETLVDCDRGATGEPIVIPTL 862

Qy        298 RRGDNDTRRF 307
              | |    ||
Db        863 GRDDGGVGRF 872


RESULT 12
S22012
erythronolide synthase (EC 2.3.1.94) III - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase III; protein A
```

C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S22012; S34784; S23206
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22012
A;Molecule type: DNA
A;Residues: 1-3172 <BEV1>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44449.1; PID:g46979
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S34784
A;Molecule type: DNA
A;Residues: 1-715,717-728,'C',729-886,888-894,'QR',895-910,'E',912-1276,'RR',1279-1284,'
EV2>
A;Cross-references: EMBL:X62569
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23206
A;Molecule type: protein
A;Residues: 2-11,'XXX',15 <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;555-835/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1117-1297/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1392-1463/Domain: acyl carrier protein homology <ACP1>
F;1511-1914/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2022-2300/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2557-2734/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2819-2890/Domain: acyl carrier protein homology <ACP2>
F;2959-3153/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 49.4%; Score 790; DB 2; Length 3172;
Best Local Similarity 56.0%; Pred. No. 8.6e-53;
Matches 169; Conservative 39; Mismatches 84; Indels 10; Gaps 5;

```
Qy          3 VFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADV 62
              |||||| || ||   ||  | |||  :: || |||| ||||: |:|   :|    :| ||
Db        557 VFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDV 613

Qy         63 VQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALA 122
              ||| ||:||||||||||:|:|| ||:||||||||||| |||:|:|||| ||:||| |
Db        614 VQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLR 673

Qy        123 AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRS-TAVSGDAEAVDEVLAYCAG 181
               : |:|||||  |  ::  : || |:|| | :|:||||||   |:|::  :||::| |
Db        674 RLGGQGGMASFGLGTEQAAERIG-RFAGALSIASVNGPRSVVVVAGESGPLDELIAECEA 732

Qy        182 TGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAY 241
              :|||||||||| | |: |||||  |  |||  : |  :||  |  :|| |:| ||
Db        733 EAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMDTAY 792

Qy        242 WYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA----IGSL 297
              || || : ||| || : ||: |   |||||||| |   ||  | |:|    |    :|:|
Db        793 WYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIE-ATLDSALPADAGACVVGTL 851

Qy        298 RR 299
              ||
Db        852 RR 853
```

RESULT 13
S13595
6-deoxyerythronolide B synthase - Saccharopolyspora erythraea
N;Alternate names: protein A
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S13595
R;Cortes, J.; Haydock, S.F.; Roberts, G.A.; Bevitt, D.J.; Leadlay, P.F.
Nature 348, 176-178, 1990
A;Title: An unusually large multifunctional polypeptide in the erythromycin-producing po
A;Reference number: S13595; MUID:91043075; PMID:2234082
A;Accession: S13595
A;Molecule type: DNA
A;Residues: 1-3178 <COR>
A;Cross-references: EMBL:X56107; NID:g46975; PIDN:CAA39583.1; PID:g46976
A;Experimental source: strain NRRL 2338
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
-carrier-protein] S-malonyltransferase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; multifunctional e
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;553-833/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1123-1303/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1398-1469/Domain: acyl carrier protein homology <ACP1>
F;1517-1920/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2028-2306/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2563-2740/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2825-2896/Domain: acyl carrier protein homology <ACP2>
F;2965-3159/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 49.4%; Score 790; DB 2; Length 3178;
Best Local Similarity 56.0%; Pred. No. 8.6e-53;
Matches 169; Conservative 39; Mismatches 84; Indels 10; Gaps 5;

```
Qy          3 VFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADV 62
              |||||| || ||   ||  | |||  :: || |||| ||||: |:|   :|    :| ||
Db        555 VFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDV 611

Qy         63 VQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALA 122
              ||| ||:||||||||||:|:|| ||:||||||||||| |||:|:|||| ||:||| |
Db        612 VQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLR 671

Qy        123 AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRS-TAVSGDAEAVDEVLAYCAG 181
               : |:|||||  |  ::  : || |:|| | :|:||||||   |:|::  :||::| |
Db        672 RLGGQGGMASFGLGTEQAAERIG-RFAGALSIASVNGPRSVVVVAGESGPLDELIAECEA 730

Qy        182 TGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAY 241
              :|||||||||| | |: |||||  |  |||  : |  :||  |  :|| |:| ||
Db        731 EAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMDTAY 790

Qy        242 WYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA----IGSL 297
              || || : ||| || : ||: |   |||||||| |   ||  | |:|    |    :|:|
Db        791 WYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIE-ATLDSALPADAGACVVGTL 849

Qy        298 RR 299
              ||
Db        850 RR 851
```

RESULT 14
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S

```
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

  Query Match             49.1%;  Score 784;  DB 2;  Length 4151;
  Best Local Similarity   50.3%;  Pred. No. 3.5e-52;
  Matches  156;  Conservative   48;  Mismatches  102;  Indels    4;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || |||: || ::| ||::: || || | :||||:||:||   |    :|
Db        559 VFVFPGQGSQWLGMGIELLDTAPAFAQQIDACAEAFAEFVDWSLVDVLRGAPGAPGLDRV 618

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||:|||||| ||:|  : ||||:||||||||||:| |||||:|||: | |||:
Db        619 DVVQPVLFAVMVSLAELWKSVAVHPDAVIGHSQGEIAAAYVAGALSLRDAARVVTLRSKL 678

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              || : | ||| |:   | :   |:   :  |: :| |||| :  |||:  |::|::| |:
Db        679 LAGLAGPGGMVSIACGADQARDLLAP-FGDRVSIAVVNGPSAVVVSGEVGALEELIAVCS 737

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              :| ||| ||||||   |: :|  | | |  | |: :   |||||  |  |||  |||
Db        738 TKELRTRRIEVDYASHSVEVEAIRGPLAEALSGIEPRSTRTVFFSTVTGNRLDTAGLDAD 797

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDT---TEDTAEDVTAIGSL 297
              |||||: | | |  ||:   : |:| |:| |||| |:  :|:|     |   :   : :|
Db        798 YWYRNVRQTVLFDQAVRNACEQGYRTFIESSPHPALITGVEETFAACTDGDSEAIVVPTL 857

Qy        298 RRGDNDTRRF 307
              |||    ||
Db        858 GRGDGGLHRF 867


RESULT 15
T03224
probable polyketide synthase module 4 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03224
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03224
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1937 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38064.1; PID:g2624951
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;549-827/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

  Query Match             48.5%;  Score 775.5;  DB 2;  Length 1937;
  Best Local Similarity   53.1%;  Pred. No. 6.4e-52;
  Matches  161;  Conservative   37;  Mismatches   96;  Indels    9;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || |||   |  || ||   :   : ||| :||||   :|  : |    :|
Db        549 VFVFPGQGSQWIGMGRGLWESSAVFRDSVLRTDRALAEFVDWSAAAVLAGEPGTPDLDRV 608

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||:|||:|| :|||||:|| ||:|||||||||||| | |||:|||: |||||||
Db        609 DVVQPVLFTVMVALANVWRSYGVEPAAVVGHSQGEIAAAHVAGGLSLRDAARVVALRSRA 668

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              | |: | |||  |       || |:  :| ||| || |||| :  |||   |::|:||
Db        669 LTALAGEGGMVVVQRSVAYVENLL-RQWEGRLSVAVVNGPEAVVVSGQVAALEELLA--- 724

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              |  ||||: ||||||   |: : |:|   | |: |  | || |||||  |:|| ::|
Db        725 -TEDRARRVAVDYASHSAQVERIEEKLTRTLTDVQPMTSRVPLFSTVERDWIDTASMDTG 783

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| | | | |||  |:: |  ||||||||| |  ::     ::  :    |:|||
Db        784 YWYRNLRQTVWFDDAVGRLSEAGFDVFVEVSPHPVLTASV----RESCPEAVVTGTLRRD 839

Qy        301 DND 303
              :|
Db        840 HDD 842


Search completed: June 17, 2003, 13:12:21
Job time : 8.85529 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 9.56552 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3_COPY_1050_1356
Perfect score: 1598
Sequence: 1 VFVFPGQGGQWAGMGVRLLA.....AEDVTAIGSLRRGDNDTRRF 307

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	891.5	55.8	3816	9	US-09-808-880-3	Sequence 3, Appli
2	861.5	53.9	4150	9	US-09-808-880-2	Sequence 2, Appli
3	858	53.7	3519	9	US-09-808-880-4	Sequence 4, Appli
4	828	51.8	344	9	US-09-735-056-33	Sequence 33, Appl
5	772.5	48.3	1346	9	US-09-860-846-37	Sequence 37, Appl
6	772.5	48.3	1346	9	US-09-988-384B-37	Sequence 37, Appl
7	772.5	48.3	1346	9	US-09-836-821-37	Sequence 37, Appl
8	772.5	48.3	1346	9	US-09-793-708-4	Sequence 4, Appli
9	772.5	48.3	1346	10	US-09-861-289-37	Sequence 37, Appl
10	772.5	48.3	3739	9	US-09-860-846-33	Sequence 33, Appl
11	772.5	48.3	3739	9	US-09-988-384B-33	Sequence 33, Appl
12	772.5	48.3	3739	9	US-09-836-821-33	Sequence 33, Appl
13	772.5	48.3	3739	9	US-09-793-708-2	Sequence 2, Appli
14	772.5	48.3	3739	10	US-09-861-289-33	Sequence 33, Appl
15	772.5	48.3	11877	9	US-09-860-846-6	Sequence 6, Appli
16	772.5	48.3	11877	9	US-09-836-821-6	Sequence 6, Appli
17	772.5	48.3	11877	10	US-09-861-289-6	Sequence 6, Appli
18	772.5	48.3	12199	9	US-09-988-384B-6	Sequence 6, Appli
19	772	48.3	345	9	US-09-735-056-34	Sequence 34, Appl
20	769	48.1	4551	9	US-09-793-708-1	Sequence 1, Appli
21	769	48.1	4613	9	US-09-860-846-31	Sequence 31, Appl
22	769	48.1	4613	9	US-09-988-384B-31	Sequence 31, Appl
23	769	48.1	4613	9	US-09-836-821-31	Sequence 31, Appl
24	769	48.1	4613	10	US-09-861-289-31	Sequence 31, Appl
25	763	47.7	1562	9	US-09-860-846-35	Sequence 35, Appl
26	763	47.7	1562	9	US-09-988-384B-35	Sequence 35, Appl
27	763	47.7	1562	9	US-09-836-821-35	Sequence 35, Appl
28	763	47.7	1562	9	US-09-793-708-3	Sequence 3, Appli
29	763	47.7	1562	10	US-09-861-289-35	Sequence 35, Appl
30	732.5	45.8	7257	9	US-10-014-717-5	Sequence 5, Appli
31	723.5	45.3	1832	9	US-10-014-717-4	Sequence 4, Appli
32	723.5	45.3	3798	9	US-10-014-717-6	Sequence 6, Appli
33	532.5	33.3	1827	9	US-09-712-363-261	Sequence 261, App
34	490	30.7	5215	9	US-09-860-846-2	Sequence 2, Appli
35	490	30.7	5215	9	US-09-988-384B-2	Sequence 2, Appli
36	490	30.7	5215	9	US-09-836-821-2	Sequence 2, Appli
37	490	30.7	5215	10	US-09-861-289-2	Sequence 2, Appli
38	440.5	27.6	1402	9	US-09-712-363-166	Sequence 166, App
39	428.5	26.8	343	9	US-09-735-056-32	Sequence 32, Appl
40	413.5	25.9	1616	9	US-09-712-363-262	Sequence 262, App
41	400.5	25.1	2439	9	US-10-014-717-7	Sequence 7, Appli
42	385.5	24.1	1421	9	US-10-014-717-2	Sequence 2, Appli
43	346	21.7	328	9	US-09-735-056-31	Sequence 31, Appl
44	317.5	19.9	2563	9	US-09-836-705-46	Sequence 46, Appl
45	317	19.8	1610	9	US-09-738-626-6666	Sequence 6666, Ap

ALIGNMENTS

```
RESULT 1
US-09-808-880-3
; Sequence 3, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

  Query Match           55.8%;  Score 891.5;  DB 9;  Length 3816;
  Best Local Similarity  58.6%;  Pred. No. 2.1e-71;
  Matches  180;  Conservative   31;  Mismatches   93;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||   || : |||| :  |   | |   ||:|::||   |:|| |
Db        567 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLR--GGEAVLGRV 624

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||  || ||:|| ||:||||||||| | | ||| | |: | |||||
Db        625 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 684
```

```
Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || ||| |  :: : : ||| ||||||| || |||||:|:||:|| |
Db        685 IARIAGGGGMVSVSLPAGRVRTML-DTYGGRLSVAAVNGPSSTVVSGDAQALDELLAGCE 743

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db        744 REGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTTALDAG 803

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| || | :    |  |:|||||
Db        804 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDTVEADAVALGSLRRD 863

Qy        301 DNDTRRF 307
              :    ||
Db        864 EGGLGRF 870


RESULT 2
US-09-808-880-2
; Sequence 2, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-2

  Query Match             53.9%;  Score 861.5;  DB 9;  Length 4150;
  Best Local Similarity   56.7%;  Pred. No. 1.2e-68;
  Matches  174;  Conservative   35;  Mismatches   95;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||   || : ||||  :  |   |     ||:|::||   |:||  |
Db       3076 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDRLTGWSLVEVLR--GGEAVLGRV 3133

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||  |::||||||  || ||:|| ||:||||||||| | | ||| | |: | |||||
Db       3134 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 3193

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || | |: |  :: : : ||: ||||||| || |||||:|:||:|| | |
Db       3194 IARIAGGGGMVSVGLSAERVRTML-DTYGGRVSVAAVNGPSSTVVSGDAQALDELLAGCE 3252

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | |::|| | |||||||   ||||| |||
Db       3253 REGVRARRVPVDYASHSAQMDQLRDELLEALADVTPQDSSVPFFSTVTADWLDTTALDAG 3312

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| || | :| :    |  |: ||||
Db       3313 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGITETLDTFDADAVALSSLRRD 3372

Qy        301 DNDTRRF 307
              :    ||
Db       3373 EGGLDRF 3379


RESULT 3
US-09-808-880-4
; Sequence 4, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 3519
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-4

  Query Match             53.7%;  Score 858;  DB 9;  Length 3519;
  Best Local Similarity   57.0%;  Pred. No. 2e-68;
  Matches  175;  Conservative   34;  Mismatches   94;  Indels    4;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR-RDAGDAVWER 59
              |||||||| || ||   || : ||||  :  |   | |   ||:|::|: |||   | |
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDA--TVLGR 2278

Qy         60 ADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
              ||||| |::||||||  || ||:|| ||:|||||||||| | | ||| | |: | ||||
Db       2279 VDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSR 2338

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              |:| : | ||| || |||  | :: : : ||: ||||||| || |||| :|:||:|| |
Db       2339 AIARIAGGGGMVSVSLPAGRVRTML-DTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGC 2397

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              |||||||:|||||||   : ||:|||| | ||:|| | |||||||   ||||| |||
Db       2398 EREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTTALDA 2457

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||: || : ||| :||: |   |   ||| |||| ||| || | :   ::   :|||||
Db       2458 GYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLRR 2517

Qy        300 GDNDTRR 306
              :    |
Db       2518 DEGGLDR 2524


RESULT 4
US-09-735-056-33
; Sequence 33, Application US/09735056
; Publication No. US20030013662A1
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
```

```
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/735,056
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. US20030013662A1e
US-09-735-056-33

  Query Match             51.8%;  Score 828;  DB 9;  Length 344;
  Best Local Similarity   55.0%;  Pred. No. 4.8e-67;
  Matches  165;  Conservative   41;  Mismatches   92;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||| |||||   ||  |  |      :|: || |:: |||: :|| :      :|
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||::||||||:||: |:|| ||:||||||||||||| ||||| |:|: |||||||
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
                 : |:||| :||:||:|:    :    |  || ||||| | | ||:|| ||: |::|
Db        125 WLGLAGKGGMVAVPMPAEELRPRL-VTWGDRLAVAAVNSPGSCAVAGDPEALAELVALLT 183

Qy        181 GTGVRARRIP-VDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              | || || || || | | | |  ||  |||:|  ::|:|: :||:||| |  || | |||
Db        184 GEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTELDA 243

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              |||||: :|| |  | :||  ||| ||:| |||| |  |:| |  |   |  :|:|||
Db        244 TYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAGTDAAVLGTLRR 303


RESULT 5
US-09-860-846-37
; Sequence 37, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-37

  Query Match             48.3%;  Score 772.5;  DB 9;  Length 1346;
  Best Local Similarity   51.4%;  Pred. No. 3.2e-61;
  Matches  160;  Conservative   42;  Mismatches   90;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| |||||  || ||| ||  |  || ||:|:||||:  ::|:       :| |
Db        562 FVFPGQGTQWAGMGAELLDSSPEFAAAMAECETALSPYVDWSLEAVVRQAPSAPTLDRVD 621

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
              ||||| |:|||||| :|: :|| |:||:|||||||||||:| |||:| |||: | |||:::
Db        622 VVQPVTFAVMVSLAKVWQHHGITPEAVIGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 681

Qy        122 AA-VRGRGGMASVPL----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 176
              || : |:||| |: |      | :| | |       | :|||||| :| ||||   : |:
Db        682 AAHLAGKGGMISLALSEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELA 735

Qy        177 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 236
                |   |:||| ||||||||  ||: :  || ::|  :|||   |||||:||||:
Db        736 QACEADGIRARIIPVDYASHSAHVETIENELADVLAGLSPQTPQVPFPSTLEGTWITEPA 795

Qy        237 LDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIG 295
              ||  ||||||   | |: ||: || |:|   |:||| || |   : |         || :
Db        796 LDGGYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPD-------KVTGLA 848

Qy        296 SLRRGDNDTRR 306
              :||| |   |
Db        849 TLRREDGGQHR 859


RESULT 6
US-09-988-384B-37
; Sequence 37, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 37
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-37

  Query Match             48.3%;  Score 772.5;  DB 9;  Length 1346;
  Best Local Similarity   51.4%;  Pred. No. 3.2e-61;
  Matches  160;  Conservative   42;  Mismatches   90;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| |||||  || ||| ||  |  || ||:|:||||:  ::|:       :| |
```

```
Db        562 FVFPGQGTQWAGMGAELLDSSPEFAAAMAECETALSPYVDWSLEAVVRQAPSAPTLDRVD 621

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121

Db        622 VVQPVTFAVMVSLAKVWQHHGITPEAVIGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 681

Qy        122 AA-VRGRGGMASVPL----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 176

Db        682 AAHLAGKGGMISLALSEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELA 735

Qy        177 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 236

Db        736 QACEADGIRARIIPVDYASHSAHVETIENELADVLAGLSPQTPQVPFFSTLEGTWITEPA 795

Qy        237 LDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIG 295

Db        796 LDGGYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPD-------KVTGLA 848

Qy        296 SLRRGDNDTRR 306

Db        849 TLRREDGGQHR 859


RESULT 7
US-09-836-821-37
; Sequence 37, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-37

  Query Match             48.3%;  Score 772.5;  DB 9;  Length 1346;
  Best Local Similarity   51.4%;  Pred. No. 3.2e-61;
  Matches  160;  Conservative   42;  Mismatches   90;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61

Db        562 FVFPGQGTQWAGMGAELLDSSPEFAAAMAECETALSPYVDWSLEAVVRQAPSAPTLDRVD 621

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121

Db        622 VVQPVTFAVMVSLAKVWQHHGITPEAVIGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 681

Qy        122 AA-VRGRGGMASVPL----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 176

Db        682 AAHLAGKGGMISLALSEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELA 735

Qy        177 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 236

Db        736 QACEADGIRARIIPVDYASHSAHVETIENELADVLAGLSPQTPQVPFFSTLEGTWITEPA 795

Qy        237 LDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIG 295

Db        796 LDGGYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPD-------KVTGLA 848

Qy        296 SLRRGDNDTRR 306

Db        849 TLRREDGGQHR 859


RESULT 8
US-09-793-708-4
; Sequence 4, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-4

  Query Match             48.3%;  Score 772.5;  DB 9;  Length 1346;
  Best Local Similarity   51.4%;  Pred. No. 3.2e-61;
  Matches  160;  Conservative   42;  Mismatches   90;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61

Db        562 FVFPGQGTQWAGMGAELLDSSPEFAAAMAECETALSPYVDWSLEAVVRQAPSAPTLDRVD 621

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121

Db        622 VVQPVTFAVMVSLAKVWQHHGITPEAVIGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 681

Qy        122 AA-VRGRGGMASVPL----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 176

Db        682 AAHLAGKGGMISLALSEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELA 735

Qy        177 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 236

Db        736 QACEADGIRARIIPVDYASHSAHVETIENELADVLAGLSPQTPQVPFFSTLEGTWITEPA 795

Qy        237 LDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIG 295

Db        796 LDGGYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPD-------KVTGLA 848

Qy        296 SLRRGDNDTRR 306

Db        849 TLRREDGGQHR 859


RESULT 9
US-09-861-289-37
; Sequence 37, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
```

```
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-37

  Query Match            48.3%;  Score 772.5;  DB 10;  Length 1346;
  Best Local Similarity   51.4%;  Pred. No. 3.2e-61;
  Matches  160;  Conservative   42;  Mismatches   90;  Indels   19;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| ||||||  || ||| ||  |  || ||:|:||||:  ::|:       :| |
Db        562 FVFPGQGTQWAGMGAELLDSSPEFAAAMAECETALSPYVDWSLEAVVRQAPSAPTLDRVD 621

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
              ||||| |:|||||| :|: :|| |:||:||||||||||:| |||:| |||: | |||::::
Db        622 VVQPVTFAVMVSLAKVWQHHGITPEAVIGHSQGEIAAAYVAGALTLDDAARVVTLRSKSI 681

Qy        122 AA-VRGRGGMASVPL----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 176
              || : |:||| |: |       | :| | |       | :|||||| :| ||||   : |:
Db        682 AAHLAGKGGMISLALSEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELA 735

Qy        177 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 236
                |   |:||| ||||||||  ||: :  || ::|  :|||   ||||||:||||:
Db        736 QACEADGIRARIIPVDYASHSAHVETIENELADVLAGLSPQTPQVPFFSTLEGTWITEPA 795

Qy        237 LDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIG 295
              ||  ||||||   | |: ||: || |:|   |:||| || |   : |         || :
Db        796 LDGGYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPD-------KVTGLA 848

Qy        296 SLRRGDNDTRR 306
              :||| |    |
Db        849 TLRREDGGQHR 859


RESULT 10
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-33

  Query Match            48.3%;  Score 772.5;  DB 9;  Length 3739;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-60;
  Matches  164;  Conservative   38;  Mismatches   92;  Indels   15;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| ||||||  ||  |  ||  |  || ||||:||||:  ::|:  |    || |
Db       2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
              ||||| |:|||||| :|: :|: | ||:|||||||||:| |||||  |||: | |||:::
Db       2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy        122 AA-VRGRGGMASVPLPAQEVEQLIGERWAG--RLWVAAVNGPRSTAVSGDAEAVDEVLAY 178
              | : |:||| |: |     |  : || ||   | |||||| :| ||||   : |:
Db       2189 GAHLAGQGGMLSLALS----EAAVVERLAGFDGLSVAAVNGPTATVVSGDPTQIQELAQA 2244

Qy        179 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 238
              |   ||||| |||||||  ||: :  || ::|  :|||   ||||||:|| |:    ||
Db       2245 CEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPALD 2304

Qy        239 AAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSL 297
              ||||||   | |: ||: || |:|   |||| || |   |: :|        || :|:|
Db       2305 GGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPET-------VTGLGTL 2357

Qy        298 RRGDNDTRR 306
              || :    |
Db       2358 RRDNGGQHR 2366


RESULT 11
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

  Query Match            48.3%;  Score 772.5;  DB 9;  Length 3739;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-60;
  Matches  164;  Conservative   38;  Mismatches   92;  Indels   15;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
              ||||||| ||||||  ||  |  ||  |  || ||||:||||:  ::|:  |    || |
Db       2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
              ||||| |:|||||| :|: :|: | ||:|||||||||:| |||||  |||: | |||:::
Db       2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy        122 AA-VRGRGGMASVPLPAQEVEQLIGERWAG--RLWVAAVNGPRSTAVSGDAEAVDEVLAY 178
              | : |:||| |: |     |  : || ||   | |||||| :| ||||   : |:
Db       2189 GAHLAGQGGMLSLALS----EAAVVERLAGFDGLSVAAVNGPTATVVSGDPTQIQELAQA 2244

Qy        179 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 238
              |   ||||| |||||||  ||: :  || ::|  :|||   ||||||:|| |:    ||
Db       2245 CEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPALD 2304

Qy        239 AAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSL 297
              ||||||   | |: ||: || |:|   |||| || |   |: :|        || :|:|
Db       2305 GGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPET-------VTGLGTL 2357

Qy        298 RRGDNDTRR 306
              || :    |
```

```
Db       2358 RRDNGGQHR 2366


RESULT 12
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             48.3%;  Score 772.5;  DB 9;  Length 3739;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-60;
  Matches  164;  Conservative   38;  Mismatches   92;  Indels   15;  Gaps    5;

Qy           2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
               ||||||| |||||| ||  |  ||  |  || ||||:||||:  ::|:  |    || |
Db        2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy          62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
               ||||| |:|||||| :|: :|: | ||:||||||||||:| ||||| |||: | |||::
Db        2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy         122 AA-VRGRGGMASVPLPAQEVEQLIGERWAG--RLWVAAVNGPRSTAVSGDAEAVDEVLAY 178
                | : |:||| |: |     |  : || ||    | ||||||| :| ||||   : |:
Db        2189 GAHLAGQGGMLSLALS----EAAVVERLAGFDGLSVAAVNGPTATVVSGDPTQIQELAQA 2244

Qy         179 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 238
               |   ||||| |||||||| ||: :  || ::|  :|||   |||||:|| |:    ||
Db        2245 CEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPALD 2304

Qy         239 AAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSL 297
                 ||||||   | |: ||: || |:|   ||||| || |  |: :|       || :|:|
Db        2305 GGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPET-------VTGLGTL 2357

Qy         298 RRGDNDTRR 306
               || :   |
Db        2358 RRDNGGQHR 2366


RESULT 13
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-2

  Query Match             48.3%;  Score 772.5;  DB 9;  Length 3739;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-60;
  Matches  164;  Conservative   38;  Mismatches   92;  Indels   15;  Gaps    5;

Qy           2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
               ||||||| |||||| ||  |  ||  |  || ||||:||||:  ::|:  |    || |
Db        2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy          62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
               ||||| |:|||||| :|: :|: | ||:||||||||||:| ||||| |||: | |||::
Db        2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy         122 AA-VRGRGGMASVPLPAQEVEQLIGERWAG--RLWVAAVNGPRSTAVSGDAEAVDEVLAY 178
                | : |:||| |: |     |  : || ||    | ||||||| :| ||||   : |:
Db        2189 GAHLAGQGGMLSLALS----EAAVVERLAGFDGLSVAAVNGPTATVVSGDPTQIQELAQA 2244

Qy         179 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 238
               |   ||||| |||||||| ||: :  || ::|  :|||   |||||:|| |:    ||
Db        2245 CEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPALD 2304

Qy         239 AAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSL 297
                 ||||||   | |: ||: || |:|   ||||| || |  |: :|       || :|:|
Db        2305 GGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPET-------VTGLGTL 2357

Qy         298 RRGDNDTRR 306
               || :   |
Db        2358 RRDNGGQHR 2366


RESULT 14
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             48.3%;  Score 772.5;  DB 10;  Length 3739;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-60;
  Matches  164;  Conservative   38;  Mismatches   92;  Indels   15;  Gaps    5;

Qy           2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
```

```
Db       2069 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 2128

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121

Db       2129 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 2188

Qy        122 AA-VRGRGGMASVPLPAQEVEQLIGERWAG--RLWVAAVNGPRSTAVSGDAEAVDEVLAY 178

Db       2189 GAHLAGQGGMLSLALS----EAAVVERLAGFDGLSVAAVNGPTATVVSGDPTQIQELAQA 2244

Qy        179 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 238

Db       2245 CEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPALD 2304

Qy        239 AAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSL 297

Db       2305 GGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPET-------VTGLGTL 2357

Qy        298 RRGDNDTRR 306

Db       2358 RRDNGGQHR 2366


RESULT 15
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6

  Query Match           48.3%;  Score 772.5;  DB 9;  Length 11877;
  Best Local Similarity  53.1%;  Pred. No. 5.7e-60;
  Matches  164;  Conservative   38;  Mismatches   92;  Indels   15;  Gaps    5;

Qy          2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61

Db       7018 FVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQAPGAPTLERVD 7077

Qy         62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121

Db       7078 VVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAARVVTLRSKSI 7137

Qy        122 AA-VRGRGGMASVPLPAQEVEQLIGERWAG--RLWVAAVNGPRSTAVSGDAEAVDEVLAY 178

Db       7138 GAHLAGQGGMLSLALS----EAAVVERLAGFDGLSVAAVNGPTATVVSGDPTQIQELAQA 7193

Qy        179 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 238

Db       7194 CEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWITEPALD 7253

Qy        239 AAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSL 297

Db       7254 GGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPET-------VTGLGTL 7306

Qy        298 RRGDNDTRR 306

Db       7307 RRDNGGQHR 7315


Search completed: June 17, 2003, 13:46:41
Job time : 10.5655 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:54:32 ; Search time 5.57989 Seconds
(without alignments)
1618.821 Million cell updates/sec

Title: US-09-914-286-3_COPY_1050_1356
Perfect score: 1598
Sequence: 1 VFVFPGQGGQWAGMGVRLLA.....AEDVTAIGSLRRGDNDTRRF 307

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	891.5	55.8	3816	4	US-09-428-517-3	Sequence 3, Appli
2	884.5	55.4	3567	2	US-07-642-734C-4	Sequence 4, Appli
3	884.5	55.4	3567	3	US-08-439-009A-4	Sequence 4, Appli
4	861.5	53.9	4150	4	US-09-428-517-2	Sequence 2, Appli
5	858	53.7	3519	4	US-09-428-517-4	Sequence 4, Appli
6	852	53.3	2595	4	US-09-036-987A-2	Sequence 2, Appli
7	852	53.3	2595	4	US-09-370-700-2	Sequence 2, Appli
8	848.5	53.1	3170	4	US-09-036-987A-4	Sequence 4, Appli
9	848.5	53.1	3170	4	US-09-370-700-4	Sequence 4, Appli
10	840.5	52.6	3491	2	US-07-642-734C-2	Sequence 2, Appli
11	840.5	52.6	3491	3	US-08-439-009A-2	Sequence 2, Appli
12	828	51.8	344	3	US-08-858-003-33	Sequence 33, Appl
13	828	51.8	344	3	US-09-078-166-33	Sequence 33, Appl
14	828	51.8	344	4	US-08-997-467-33	Sequence 33, Appl
15	822.5	51.5	3724	2	US-08-804-227C-10	Sequence 10, Appl
16	822.5	51.5	3724	2	US-08-804-198-4	Sequence 4, Appli
17	816.5	51.1	829	4	US-09-413-814-105	Sequence 105, App
18	814.5	51.0	3170	2	US-07-642-734C-5	Sequence 5, Appli
19	814.5	51.0	3170	3	US-08-439-009A-5	Sequence 5, Appli
20	774	48.4	5588	4	US-09-036-987A-6	Sequence 6, Appli
21	774	48.4	5588	4	US-09-370-700-6	Sequence 6, Appli
22	773	48.4	1580	2	US-08-804-227C-11	Sequence 11, Appl
23	773	48.4	1580	2	US-08-804-198-5	Sequence 5, Appli
24	772.5	48.3	1346	3	US-09-320-878-4	Sequence 4, Appli
25	772.5	48.3	1346	4	US-09-105-537-37	Sequence 37, Appl
26	772.5	48.3	3739	3	US-09-320-878-2	Sequence 2, Appli
27	772.5	48.3	3739	4	US-09-105-537-33	Sequence 33, Appl
28	772.5	48.3	11877	4	US-09-105-537-6	Sequence 6, Appli
29	772	48.3	345	3	US-08-858-003-34	Sequence 34, Appl
30	772	48.3	345	3	US-09-078-166-34	Sequence 34, Appl
31	772	48.3	345	4	US-08-997-467-34	Sequence 34, Appl
32	769	48.1	4551	3	US-09-320-878-1	Sequence 1, Appli
33	769	48.1	4613	4	US-09-105-537-31	Sequence 31, Appl
34	763	47.7	1562	3	US-09-320-878-3	Sequence 3, Appli
35	763	47.7	1562	4	US-09-105-537-35	Sequence 35, Appl
36	754	47.2	3729	2	US-08-804-227C-4	Sequence 4, Appli
37	736	46.1	4472	2	US-08-804-227C-2	Sequence 2, Appli
38	733	45.9	1611	2	US-08-804-227C-5	Sequence 5, Appli
39	733	45.9	4545	2	US-08-804-227C-14	Sequence 14, Appl
40	732.5	45.8	7257	3	US-09-335-409-5	Sequence 5, Appli
41	732.5	45.8	7257	4	US-09-568-102-5	Sequence 5, Appli
42	732.5	45.8	7257	4	US-09-567-969-5	Sequence 5, Appli
43	732.5	45.8	7257	4	US-09-568-480-5	Sequence 5, Appli
44	732.5	45.8	7257	4	US-09-568-486-5	Sequence 5, Appli
45	732.5	45.8	7257	4	US-09-568-472-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3

Query Match 55.8%; Score 891.5; DB 4; Length 3816;
Best Local Similarity 58.6%; Pred. No. 5.7e-81;
Matches 180; Conservative 31; Mismatches 93; Indels 3; Gaps 2;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||| || ||   || : |||| : |  | | ||:|::||  |:||  |
Db        567 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLR--GGEAVLGRV 624

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||  || ||:|| ||:||||||||| | | ||| | |: | |||||
Db        625 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 684

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || ||| |  :: : : ||| |||||||| || ||||||:|:||:|| |
Db        685 IARIAGGGGMVSVSLPAGRVRTML-DTYGGRLSVAAVNGPSSTVVSGDAQALDELLAGCE 743

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db        744 REGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTTALDAG 803

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| || | :    |  |:|||||
```

```
Db        804 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDTVEADAVALGSLRRD 863

Qy        301 DNDTRRF 307
              :   ||
Db        864 EGGLGRF 870
```

RESULT 2
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Danckers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-642-734C-4

Query Match 55.4%; Score 884.5; DB 2; Length 3567;
Best Local Similarity 57.3%; Pred. No. 2.6e-80;
Matches 177; Conservative 41; Mismatches 88; Indels 3; Gaps 2;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||:|||| |||||| ||:|||||| :::||:|::||  || | |:||:  |    :|
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||:|||||| ||||||:|| ||:||||||||||||| |||:|:|||| |  |||
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : :: | ||||:| |    | : :   |  || |||||||||  |||:  |:       ||
Db        680 MRSLSGEGGMAAVALGEAAVRERL-RPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCA 738

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                |:| | | |||||| | :: :|||||  |||:|:|: | | ||||   :| | |||
Db        739 AEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTELDAR 798

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTE--DTAEDVTAIGSLR 298
              |||||| : |||:|||  ||: |:  |:|||||| :| |:|:  |  | |||   :|||
Db        799 YWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAEDAVVVGSLH 858
```

```
Qy        299 RGDNDTRRF 307
                |   |   |
Db        859 RDGGDLSAF 867
```

RESULT 3
US-08-439-009A-4
; Sequence 4, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-439-009A-4

Query Match 55.4%; Score 884.5; DB 3; Length 3567;
Best Local Similarity 57.3%; Pred. No. 2.6e-80;
Matches 177; Conservative 41; Mismatches 88; Indels 3; Gaps 2;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||:|||| |||||| ||:|||||| :::||:|::||  || | |:||:  |    :|
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||:|||||| ||||||:|| ||:||||||||||||| |||:|:|||| |  |||
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : :: | ||||:| |    | : :   |  || |||||||||  |||:  |:       ||
Db        680 MRSLSGEGGMAAVALGEAAVRERL-RPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCA 738

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                |:| | | |||||| | :: :|||||  |||:|:|: | | ||||   :| | |||
Db        739 AEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTELDAR 798

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTE--DTAEDVTAIGSLR 298
              |||||| : |||:|||  ||: |:  |:|||||| :| |:|:  |  | |||   :|||
Db        799 YWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAEDAVVVGSLH 858

Qy        299 RGDNDTRRF 307
```

```
                |   |   |
Db        859 RDGGDLSAF 867


RESULT 4
US-09-428-517-2
; Sequence 2, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4150
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-2

  Query Match             53.9%;  Score 861.5;  DB 4;  Length 4150;
  Best Local Similarity   56.7%;  Pred. No. 7.3e-78;
  Matches  174;  Conservative   35;  Mismatches   95;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||   || : |||| :  |   |     ||:|::||   |:|| |
Db       3076 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDRLTGWSLVEVLR--GGEAVLGRV 3133

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::||||||   || ||:|| ||:||||||||||| | | ||| | |: | |||||
Db       3134 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 3193

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || | |: |  :: : : ||: ||||||| || |||||:|:||:|| |
Db       3194 IARIAGGGGMVSVGLSAERVRTML-DTYGGRVSVAAVNGPSSTVVSGDAQALDELLAGCE 3252

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | |::|| | |||||||   ||||| |||
Db       3253 REGVRARRVPVDYASHSAQMDQLRDELLEALADVTPQDSSVPFFSTVTADWLDTTALDAG 3312

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| | :| :    |  |: ||||
Db       3313 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGITETLDTFDADAVALSSLRRD 3372

Qy        301 DNDTRRF 307
              :   ||
Db       3373 EGGLDRF 3379


RESULT 5
US-09-428-517-4
; Sequence 4, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 3519
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-4

  Query Match             53.7%;  Score 858;  DB 4;  Length 3519;
  Best Local Similarity   57.0%;  Pred. No. 1.3e-77;
  Matches  175;  Conservative   34;  Mismatches   94;  Indels    4;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR-RDAGDAVWER 59
              |||||||| || ||   || : |||| :  |   | |   ||:|::|: |||   | |
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDA--TVLGR 2278

Qy         60 ADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
              ||||| |::||||||   || ||:|| ||:|||||||||| | | ||| | |: | ||||
Db       2279 VDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSR 2338

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              |:| : | ||| || |||  |  :: : : ||: ||||||| || |||| :|:||:|| |
Db       2339 AIARIAGGGGMVSVSLPAGRVRTML-DTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGC 2397

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db       2398 EREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTTALDA 2457

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||: || : ||| :||: |   |   ||| |||| ||| || | :   ::   :|||||
Db       2458 GYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLRR 2517

Qy        300 GDNDTRR 306
              :   |
Db       2518 DEGGLDR 2524


RESULT 6
US-09-036-987A-2
; Sequence 2, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2595 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-2

  Query Match             53.3%;  Score 852;  DB 4;  Length 2595;
  Best Local Similarity   55.2%;  Pred. No. 3.3e-77;
  Matches  169;  Conservative   47;  Mismatches   88;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||| :|| :| |||  |  | :|||| :|||::|:||  || |  :
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| ||::||||| ||||:|: | ||:||||||||| | ||||::|||: ||:|||
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              | |: | | |||:  ||:|| |::   |  |: || |||| || |||| ||: |:|| ||
Db        647 LTALAGSGAMASLQHPAEEVRQIL-LPWRDRIGVAGVNGPSSTLVSGDREAMAELLAECA 705

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :| |||||:|||| ||:: :|:||| ||  : |:   :| :||  |  ||   :||
Db        706 DRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPMDAD 764

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||| ||| |  ||:||   |:  |:|:|||| |   |::|      :| |:|:|||
Db        765 YWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAGREVVALGTLRRG 824

Qy        301 DNDTRR 306
              :   |:
Db        825 EGGMRQ 830


RESULT 7
US-09-370-700-2
; Sequence 2, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2595
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-2

  Query Match             53.3%;  Score 852;  DB 4;  Length 2595;
  Best Local Similarity   55.2%;  Pred. No. 3.3e-77;
  Matches  169;  Conservative   47;  Mismatches   88;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||| :|| :| |||  |  | :|||| :|||::|:||  || |  :
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| ||::||||| ||||:|: | ||:||||||||| | ||||::|||: ||:|||
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              | |: | | |||:  ||:|| |::   |  |: || |||| || |||| ||: |:|| ||
Db        647 LTALAGSGAMASLQHPAEEVRQIL-LPWRDRIGVAGVNGPSSTLVSGDREAMAELLAECA 705

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :| |||||:|||| ||:: :|:||| ||  : |:   :| :||  |  ||   :||
Db        706 DRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPMDAD 764

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||| ||| |  ||:||   |:  |:|:|||| |   |::|      :| |:|:|||
Db        765 YWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAGREVVALGTLRRG 824

Qy        301 DNDTRR 306
              :   |:
Db        825 EGGMRQ 830


RESULT 8
US-09-036-987A-4
; Sequence 4, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3170 amino acids
;      TYPE:  amino acid
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-4

  Query Match             53.1%;  Score 848.5;  DB 4;  Length 3170;
  Best Local Similarity   56.2%;  Pred. No. 1e-76;
  Matches  172;  Conservative   35;  Mismatches   98;  Indels    1;  Gaps    1;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| ||||| | || :|| ||| |   |  ||  :||||:||:||        :|
Db        563 VFVFPGQGWQWAGMAVDLLDASPTFARHMDECATALRRYVDWSLVDVLRGAENSPPLDRV 622

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:|||||| :|||||: | ||:||||||||||   | | |:|||: |||||||
Db        623 DVLQPASFAVMVSLAEVWRSYGVRPAAVVGHSQGEIAAACAAGVLPLEDAARLVALRSRA 682

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |  : |||||||:  || ||  |      ||| |||:|||||  |||| |||||:|| ||
Db        683 LKGLSGRGGMASLACPADEVAALFAGS-GGRLEVAAINGPRSVVVSGDLEAVDELLAECA 741

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              :|||||||||||  ||: :|  :|     :  :   ||::||| | |:|    ||
Db        742 EKDMRARRIPVDYASHSAHVEVVRSPVLAAAAGVRHRDGQVPWWSTVIGDWVDPARLDGE 801

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| |||||  ||| | : |  :|:|:| || |  |:|:|  ::    | |:|:|||
Db        802 YWYRNLRQPVRFEHAVQGLVERGFGLFIEMSAHPVLTTAVEETGAESETAVAAVGTLRRD 861

Qy        301 DNDTRR 306
                 ||
Db        862 SGGLRR 867


RESULT 9
US-09-370-700-4
; Sequence 4, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 3170
;    TYPE: PRT
;    ORGANISM: Saccharopolyspora spinosa
US-09-370-700-4

  Query Match             53.1%;  Score 848.5;  DB 4;  Length 3170;
  Best Local Similarity   56.2%;  Pred. No. 1e-76;
  Matches  172;  Conservative   35;  Mismatches   98;  Indels    1;  Gaps    1;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| ||||| | || :|| ||| |   |  ||  :||||:||:||        :|
Db        563 VFVFPGQGWQWAGMAVDLLDASPTFARHMDECATALRRYVDWSLVDVLRGAENSPPLDRV 622

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:|||||| :|||||: | ||:||||||||||   | | |:|||: |||||||
Db        623 DVLQPASFAVMVSLAEVWRSYGVRPAAVVGHSQGEIAAACAAGVLPLEDAARLVALRSRA 682

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |  : |||||||:  || ||  |      ||| |||:|||||  |||| |||||:|| ||
Db        683 LKGLSGRGGMASLACPADEVAALFAGS-GGRLEVAAINGPRSVVVSGDLEAVDELLAECA 741

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              :|||||||||||  ||: :|  :|     :  :   ||::||| | |:|    ||
Db        742 EKDMRARRIPVDYASHSAHVEVVRSPVLAAAAGVRHRDGQVPWWSTVIGDWVDPARLDGE 801

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| |||||  ||| | : |  :|:|:| || |  |:|:|  ::    | |:|:|||
Db        802 YWYRNLRQPVRFEHAVQGLVERGFGLFIEMSAHPVLTTAVEETGAESETAVAAVGTLRRD 861

Qy        301 DNDTRR 306
                 ||
Db        862 SGGLRR 867


RESULT 10
US-07-642-734C-2
; Sequence 2, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3491 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-642-734C-2

  Query Match             52.6%;  Score 840.5;  DB 2;  Length 3491;
  Best Local Similarity   55.1%;  Pred. No. 7.7e-76;
  Matches  173;  Conservative   46;  Mismatches   84;  Indels   11;  Gaps    4;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAG----DAV 56
              |||||||| ||||| | || :|||||  :: | :|| | :|: |:  || :|     ||
Db       1031 VFVFPGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAA 1090

Qy         57 --WERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTV 114
                || ||||||:|:||||||::||::|:|| ||:|||||||||| | ||||| |||: |
Db       1091 LSTERVDVVQPVMFAVMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVV 1150

Qy        115 ALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDE 174
              ||||| :| : |  ||||:  || ||   ||:    |: :||||||||  |:||:: :|
Db       1151 ALRSRVIATMPGNKGMASIAAPAGEVRARIGD----RVEIAAVNGPRSVVVAGDSDELDR 1206
```

```
Qy        175 VLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLG-DISPQPSGVPFFSTVEGTWLD 233
              ::| |    :||:|: ||||||  ||: :|: |   || |  | |  ||||||| | |
Db       1207 LVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDFHPLPGFVPFFSTVTGRWTQ 1266

Qy        234 TTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA 293
              ||| |||||| : |||:|||:|||: |:| |:||| || |  |||:  : :  |::|
Db       1267 PDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGDGSGADLSA 1326

Qy        294 IGSLRRGDNDTRRF 307
              | ||||||     |
Db       1327 IHSLRRGDGSLADF 1340


RESULT 11
US-08-439-009A-2
; Sequence 2, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3491 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-439-009A-2

  Query Match             52.6%;  Score 840.5;  DB 3;  Length 3491;
  Best Local Similarity   55.1%;  Pred. No. 7.7e-76;
  Matches  173;  Conservative   46;  Mismatches   84;  Indels   11;  Gaps    4;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAG----DAV 56
              |||||||| ||||| | || :|||||  :: | :|| | :|: |:  || :|      ||
Db       1031 VFVFPGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAA 1090

Qy         57 --WERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTV 114
                || |||||:|:||||||::||::|:|| ||:||||||||| | ||||| |||: |
Db       1091 LSTERVDVVQPVMFAVMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVV 1150

Qy        115 ALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDE 174
              ||||| :| : |  ||||:  || ||   ||:    |: :||||||||  |:||:: :|
Db       1151 ALRSRVIATMPGNKGMASIAAPAGEVRARIGD----RVEIAAVNGPRSVVVAGDSDELDR 1206

Qy        175 VLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLG-DISPQPSGVPFFSTVEGTWLD 233
              ::| |    :||:|: ||||||  ||: :|: |   || |  | |  ||||||| | |
Db       1207 LVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDFHPLPGFVPFFSTVTGRWTQ 1266

Qy        234 TTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTA 293
              ||| |||||| : |||:|||:|||: |:| |:||| || |  |||:  : :  |::|
Db       1267 PDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGDGSGADLSA 1326

Qy        294 IGSLRRGDNDTRRF 307
              | ||||||     |
Db       1327 IHSLRRGDGSLADF 1340


RESULT 12
US-08-858-003-33
; Sequence 33, Application US/08858003
; Patent No. 6060234
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,003
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6060234e
US-08-858-003-33

  Query Match             51.8%;  Score 828;  DB 3;  Length 344;
  Best Local Similarity   55.0%;  Pred. No. 4.3e-76;
  Matches  165;  Conservative   41;  Mismatches   92;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| |||||   ||  |  |      :|: || |:: ||| : :|| :      :|
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||::||||||:||: |:|| ||:|||||||||||| ||||| |:|: |||||||
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124
```

```
Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : |:||| :||:||:|:   :    |  || ||||| | | ||:|| ||: |::|
Db        125 WLGLAGKGGMVAVPMPAEELRPRL-VTWGDRLAVAAVNSPGSCAVAGDPEALAELVALLT 183

Qy        181 GTGVRARRIP-VDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              | || || || || | | | |  || |||:|  ::|:|: :||:||| |  || | |||
Db        184 GEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTELDA 243

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              |||||: :|| |  | :||  ||| ||:| |||| |  |:| |  |   |   :|:|||
Db        244 TYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAGTDAAVLGTLRR 303


RESULT 13
US-09-078-166-33
; Sequence 33, Application US/09078166
; Patent No. 6063561
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,166
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6063561e
US-09-078-166-33

  Query Match           51.8%;  Score 828;  DB 3;  Length 344;
  Best Local Similarity   55.0%;  Pred. No. 4.3e-76;
  Matches  165;  Conservative   41;  Mismatches   92;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| |||||   ||  |  |     :|: || |:: |||: :|| :      :|
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||::||||||:||: |:|| ||:|||||||||||| ||||| |:|: |||||||
Db         65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : |:||| :||:||:|:   :    |  || ||||| | | ||:|| ||: |::|
Db        125 WLGLAGKGGMVAVPMPAEELRPRL-VTWGDRLAVAAVNSPGSCAVAGDPEALAELVALLT 183

Qy        181 GTGVRARRIP-VDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              | || || || || | | | |  || |||:|  ::|:|: :||:||| |  || | |||
Db        184 GEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTELDA 243

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              |||||: :|| |  | :||  ||| ||:| |||| |  |:| |  |   |   :|:|||
Db        244 TYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAGTDAAVLGTLRR 303


RESULT 14
US-08-997-467-33
; Sequence 33, Application US/08997467
; Patent No. 6200813
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,467
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:  16-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6200813e
US-08-997-467-33

  Query Match           51.8%;  Score 828;  DB 4;  Length 344;
  Best Local Similarity   55.0%;  Pred. No. 4.3e-76;
  Matches  165;  Conservative   41;  Mismatches   92;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| |||||   ||  |  |     :|: || |:: |||: :|| :      :|
Db          5 VFVFPGQGSQWAGMAEGLLERSGAFRSAADSCDAALRPYLGWSVLSVLRGEPDAPSLDRV 64
```

```
Qy          61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
               ||||||||::||||||:||: |:|| ||:|||||||||||| ||||| |:|: |||||||
Db          65 DVVQPVLFTMMVSLAAVWRALGVEPAAVVGHSQGEIAAAHVAGALSLDDSARIVALRSRA 124

Qy         121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
                 : |:||| :||:||:|:  :    |  || ||||| | | ||:|| ||: |::|
Db         125 WLGLAGKGGMVAVPMPAEELRPRL-VTWGDRLAVAAVNSPGSCAVAGDPEALAELVALLT 183

Qy         181 GTGVRARRIP-VDYASHCPHVQPLREELLELLGLDISPQPSGVPFFSTVEGTWLDTTTLDA 239
               | || || || || | | | |  ||  |||:|  ::|:|: :||:||| |  || | |||
Db         184 GEGVHARPIPGVDTAGHSPQVDALRAHLLEVLAPVAPRPADIPFYSTVTGGLLDGTELDA 243

Qy         240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
                |||||: :|| |  | :||  ||| ||:| |||| |  |:| |  |   |   :|:|||
Db         244 TYWYRNMREPVEFERATRALIADGHDVFLETSPHPMLAVALEQTVTDAGTDAAVLGTLRR 303


RESULT 15
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3724 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-10

  Query Match           51.5%;  Score 822.5;  DB 2;  Length 3724;
  Best Local Similarity  55.0%;  Pred. No. 5.7e-74;
  Matches  169;  Conservative   36;  Mismatches   97;  Indels    5;  Gaps    2;

Qy           2 FVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERAD 61
               ||||||| |||||| || ::| ||  :  |: ||:|:|||:: |:||        :| |
Db         594 FVFPGQGSQWAGMGAELLDTAPAFAAELDRCQGALSPYVDWNLADVLRGAPAAPGLDRVD 653

Qy          62 VVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRAL 121
               ||||  |:||| ||||||| |:|| ||:|||||||||| | |||||:|||: |||||: :
Db         654 VVQPATFAVMVGLAALWRSLGVEPAAVIGHSQGEIAAACVAGALSLEDAARIVALRSQVI 713

Qy         122 A-AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
               |  : |||||||| ||| |||     | || : :|||||| || | |:  |:: :|

Db         714 ARELAGRGGMASVALPAAEVE----ARLAGGVEIAAVNGPGSTVVCGEPGALEALLVTLE 769

Qy         181 GTGVRARRIPVDYASHCPHVQPLREELLELLGLDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                 | | ||| ||||||  :|: :| ||  :|| : |:   |||:||||   ||| ||||
Db         770 SEGTRVRRIDVDYASHSHYVESIRAELATVLGPVRPRRGDVPFYSTVEAALLDTATLDAD 829

Qy         241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
               |||||| ||||   |:|: |||   ||| | || |   :  | |     | |: ||||
Db         830 YWYRNLRLPVRFEPTVRAMLDDGVDAFVECSAHPVLTVGVRQTVESAGGAVPALASLRRD 889

Qy         301 DNDTRRF 307
               :  |||
Db         890 EGGLRRF 896


Search completed: June 17, 2003, 13:09:26
Job time : 6.57989 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:50:57 ; Search time 15.0391 Seconds
(without alignments)
2720.102 Million cell updates/sec

Title: US-09-914-286-3_COPY_1050_1356
Perfect score: 1598
Sequence: 1 VFVFPGQGGQWAGMGVRLLA.....AEDVTAIGSLRRGDNDTRRF 307

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1598	100.0	3972	21	AAB23749	S. avermitilis ave
2	1598	100.0	3972	22	AAG65264	Streptomyces averm
3	1598	100.0	3972	22	AAG65268	Streptomyces averm
4	1598	100.0	5532	21	AAB23752	S. avermitilis ave
5	1598	100.0	5532	22	AAG65267	Streptomyces averm
6	1366	85.5	6239	21	AAB23750	S. avermitilis ave
7	1366	85.5	6239	22	AAG65265	Streptomyces averm
8	891.5	55.8	3816	21	AAY92708	S. antibioticus 8,
9	884.5	55.4	3567	14	AAR44431	eryA region polype
10	861.5	53.9	4150	21	AAY92707	S. antibioticus 8,
11	858	53.7	3519	21	AAY92709	S. antibioticus 8,
12	852	53.3	2595	20	AAY39297	SpnA a polyketide
13	852	53.3	2595	22	AAB70965	S. spinosa protein
14	848.5	53.1	3170	20	AAY39299	SpnC a polyketide
15	848.5	53.1	3170	22	AAB70967	S. spinosa protein
16	848.5	53.1	5069	19	AAW52846	A. mediterranei ri
17	840.5	52.6	3398	14	AAR44430	eryA region polype
18	832	52.1	3546	22	AAB82212	Polyketide synthas
19	832	52.1	4881	21	AAB23751	S. avermitilis ave
20	832	52.1	4881	22	AAG65266	Streptomyces averm
21	828.5	51.8	3192	22	AAE10128	Streptomyces nours
22	828	51.8	344	20	AAW87714	An active acyltran
23	828	51.8	344	21	AAB12926	Protein sequence o
24	827	51.8	3562	22	AAB82213	Polyketide synthas
25	826	51.7	7068	22	AAE10142	Streptomyces nours
26	826	51.7	9477	22	AAE10144	Streptomyces nours
27	822.5	51.5	3724	18	AAW23718	Platenolide syntha
28	822.5	51.5	3724	18	AAW22608	Platenolide syntha
29	815.5	51.0	4572	19	AAW52845	A. mediterranei ri
30	814.5	51.0	2986	14	AAR44432	eryA region polype
31	812.5	50.8	1721	19	AAW52847	A. mediterranei ri
32	809.5	50.7	3201	22	AAB82214	Polyketide synthas
33	806.5	50.5	3413	19	AAW52849	A. mediterranei ri
34	805	50.4	1688	19	AAW52848	A. mediterranei ri
35	774	48.4	5588	20	AAY39301	SpnE a polyketide
36	774	48.4	5588	22	AAB70969	S. spinosa protein
37	773	48.4	1580	18	AAW23719	Platenolide syntha
38	773	48.4	1580	18	AAW22609	Platenolide syntha
39	772.5	48.3	1346	21	AAB18640	Amino acid sequenc
40	772.5	48.3	1346	21	AAY77195	S. venezuelae macr
41	772.5	48.3	1346	21	AAY67204	Narbonolide syntha
42	772.5	48.3	3739	21	AAB18638	Amino acid sequenc
43	772.5	48.3	3739	21	AAY77193	S. venezuelae macr
44	772.5	48.3	3739	21	AAY77201	S. venezuelae pik
45	772.5	48.3	3739	21	AAY67202	Narbonolide syntha

ALIGNMENTS

RESULT 1
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 1598; DB 21; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2.1e-148;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
Db 1050 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 1109

Qy 61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
Db 1110 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 1169

Qy 121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
Db 1170 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 1229

Qy 181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
Db 1230 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 1289

Qy 241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
Db 1290 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 1349

Qy 301 DNDTRRF 307
Db 1350 DNDTRRF 1356

RESULT 2
AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 1598; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 2.1e-148;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
Db 1050 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 1109

Qy 61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
Db 1110 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 1169

Qy 121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
Db 1170 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 1229

Qy 181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
Db 1230 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 1289

Qy 241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
Db 1290 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 1349

Qy 301 DNDTRRF 307
Db 1350 DNDTRRF 1356

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.

```
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79279.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 5; Page 235-248; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is derived from an S.
CC   avermitilis protein.
XX
SQ   Sequence   3972 AA;

  Query Match            100.0%;  Score 1598;  DB 22;  Length 3972;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-148;
  Matches  307;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 1109

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 1169

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 1229

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 1289

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 1349

Qy        301 DNDTRRF 307
              |||||||
Db       1350 DNDTRRF 1356


RESULT 4
AAB23752
ID   AAB23752 standard; Protein; 5532 AA.
XX
AC   AAB23752;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 278-305; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;

  Query Match            100.0%;  Score 1598;  DB 21;  Length 5532;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-148;
  Matches  307;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 662

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 722

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 782

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 842

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 902

Qy        301 DNDTRRF 307
              |||||||
Db        903 DNDTRRF 909


RESULT 5
AAG65267
ID   AAG65267 standard; Protein; 5532 AA.
XX
AC   AAG65267;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 7.
XX
```

KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79278.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 217-235; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 5532 AA;

Query Match 100.0%; Score 1598; DB 22; Length 5532;
Best Local Similarity 100.0%; Pred. No. 3.5e-148;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 662

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 722

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 782

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 842

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 902

Qy        301 DNDTRRF 307
              |||||||
Db        903 DNDTRRF 909
```

RESULT 6
AAB23750
ID AAB23750 standard; Protein; 6239 AA.
XX
AC AAB23750;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 32; Page 223-254; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 6239 AA;

Query Match 85.5%; Score 1366; DB 21; Length 6239;
Best Local Similarity 81.8%; Pred. No. 4.3e-125;
Matches 251; Conservative 25; Mismatches 31; Indels 0; Gaps 0;

```
Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||||||||||: || ||||||  : |||:|| ||| ||: ||| ||  |  |::|
Db       4815 VFVFPGQGGQWAGMGLTLLTSSPVFAEHIDACEKALTPWVPWSLTDILHRDPDDPAWQQA 4874

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||||||:||||||||||||||||||||||||||||:||||||||||||||||||||
Db       4875 DVVQPVLFSIMVSLAALWRSYGIEPDAVLGHSQGEIAAAHICGALSLKDAAKTVALRSRA 4934

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              ||||||||| |||:|||||:|:||| ||| |:|||||:||| || |||| :|||||||:|
Db       4935 LAAVRGRGAMASLPLPAQDVQQLISERWEGQLWVAALNGPHSTTVSGDTKAVDEVLAHCT 4994

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||:||:|||||||||||||||| :||| |||||:|||| ||||||||||||||||||||
Db       4995 DTGLRAKRIPVDYASHCPHVQPLHDELLHLLGDITPQPSTVPFFSTVEGTWLDTTTLDAA 5054

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||||||||| |:| | ||||| |:|:|||||||||||||:| |::|| ||||||
Db       5055 YWYRNLHQPVRFSHAIQTLTDDGHRAFIEISPHPTLVPAIEDTTENTTENITATGSLRRG 5114

Qy        301 DNDTRRF 307
              |||| ||
Db       5115 DNDTHRF 5121
```

```
RESULT 7
AAG65265
ID   AAG65265 standard; Protein; 6239 AA.
XX
AC   AAG65265;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis protein SEQ ID NO: 5.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   N-PSDB; AAH79277.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 4; Page 180-201; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ   Sequence   6239 AA;

  Query Match             85.5%;  Score 1366;  DB 22;  Length 6239;
  Best Local Similarity   81.8%;  Pred. No. 4.3e-125;
  Matches  251;  Conservative   25;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||||||||||: || ||||||  : |||:|| ||| ||: ||| ||  |  |::|
Db       4815 VFVFPGQGGQWAGMGLTLLTSSPVFAEHIDACEKALTPWVPWSLTDILHRDPDDPAWQQA 4874

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              |||||||||:|||||||||||||||||||||||||||||:||||||||||||||||||
Db       4875 DVVQPVLFSIMVSLAALWRSYGIEPDAVLGHSQGEIAAAHICGALSLKDAAKTVALRSRA 4934

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |||||||| |||:|||||:|:||| ||| |:|||||:||| || |||| :|||||||:|
Db       4935 LAAVRGRGAMASLPLPAQDVQQLISERWEGQLWVAALNGPHSTTVSGDTKAVDEVLAHCT 4994

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||:||:|||||||||||||| :||| |||||:|||| ||||||||||||||||||||
Db       4995 DTGLRAKRIPVDYASHCPHVQPLHDELLHLLGDITPQPSTVPFFSTVEGTWLDTTTLDAA 5054

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||||||||| |:| | ||||| |:|:|||||||||||||||:| |::|| ||||||
Db       5055 YWYRNLHQPVRFSHAIQTLTDDGHRAFIEISPHPTLVPAIEDTTENTTENITATGSLRRG 5114

Qy        301 DNDTRRF 307
              |||| ||
Db       5115 DNDTHRF 5121


RESULT 8
AAY92708
ID   AAY92708 standard; Protein; 3816 AA.
XX
AC   AAY92708;
XX
DT   29-AUG-2000  (first entry)
XX
DE   S. antibioticus 8,8a-deoxyoleandolide synthase II encoded by oleAII.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   Region          30..1573
FT                   /label= Extender_module_3
FT   Domain          30..459
FT                   /label= KS3_domain
FT                   /note= "ketosynthase domain"
FT   Domain          566..911
FT                   /label= AT3_domain
FT                   /note= "acyl transferase domain"
FT   Domain          1210..1394
FT                   /label= KR3_domain
FT                   /note= "ketoreductase domain"
FT   Domain          1487..1573
FT                   /label= ACP3_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          1593..3710
FT                   /label= Extender_module_4
FT   Domain          1593..2021
FT                   /label= KS4_domain
FT                   /note= "ketosynthase domain"
FT   Domain          2126..2472
FT                   /label= AT4_domain
FT                   /note= "acyl transferase domain"
FT   Domain          2484..2663
FT                   /label= DH4_domain
FT                   /note= "dehydratase domain"
FT   Domain          3054..3345
FT                   /label= ER4_domain
FT                   /note= "enoyl reductase domain"
FT   Domain          3349..3532
FT                   /label= KR4_domain
FT                   /note= "ketoreductase domain"
FT   Domain          3627..3710
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein domain"
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24478.
XX
PR   29-OCT-1998;   98US-0106100.
PR   16-FEB-1999;   99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   N-PSDB; AAA09469.
```

```
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 28-29; 86pp; English.
XX
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence   3816 AA;

  Query Match             55.8%;  Score 891.5;  DB 21;  Length 3816;
  Best Local Similarity   58.6%;  Pred. No. 2.6e-78;
  Matches  180;  Conservative   31;  Mismatches   93;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||   || : ||||  :  |   | |   ||:|::||   |:|| |
Db        567 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLR--GGEAVLGRV 624

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||| .|| ||:|| ||:||||||||| | | ||| | |: | |||||
Db        625 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 684

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || |||  |  :: : : ||| ||||||| || |||||:|:||:|| |
Db        685 IARIAGGGGMVSVSLPAGRVRTML-DTYGGRLSVAAVNGPSSTVVSGDAQALDELLAGCE 743

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db        744 REGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTTALDAG 803

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |   |   ||| |||| ||| || | :    |  |:|||||
Db        804 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDTVEADAVALGSLRRD 863

Qy        301 DNDTRRF 307
              :   ||
Db        864 EGGLGRF 870


RESULT 9
AAR44431
ID   AAR44431 standard; Protein; 3567 AA.
XX
AC   AAR44431;
XX
DT   22-DEC-1993  (first entry)
XX
DE   eryA region polypeptide module #2.
XX
KW   Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW   erythromycin; condensation; elongation; acyl chain growth;
KW   gene replacement.
XX
OS   Saccarapolyspora erythraea.
XX
PN   WO9313663-A.
XX
PD   22-JUL-1993.
XX
PF   17-JAN-1992;   92WO-US00427.
XX
PR   17-JAN-1992;   92WO-US00427.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Donadio S,  Katz L,  McAlpine JB;
XX
DR   WPI; 1993-242804/30.
DR   N-PSDB; AAQ46806.
XX
PT   Biosynthesis of specific polyketide analogues esp. erythromycin
PT   cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT   microorganisms
XX
PS   Disclosure; Fig 2; 133pp; English.
XX
CC   The sequences given in AAR44430-32 are encoded by the eryA fragment of
CC   the Saccarapolyspora erythraea genome.  These polypeptides are
CC   involved in the biosynthesis of the polyketide segment of erythromycin.
CC   eryA is organised in modules and each module takes care of one
CC   condensation step.  The precise succession of elongation steps is
CC   dictated by the genetic order of the modules.  The DNA encoding
CC   these polypeptides may be specifically altered such that novel
CC   polyketide molecules of desired structure are produced.  Three types
CC   of alteration may be produced; those inactivating a single function in
CC   a module which does not arrest acyl chain growth; those inactivating a
CC   single function in a module which does affect chain growth; and those
CC   affecting an entire module.  The mutations may be introduced by gene
CC   replacement.
XX
SQ   Sequence   3567 AA;

  Query Match             55.4%;  Score 884.5;  DB 14;  Length 3567;
  Best Local Similarity   57.3%;  Pred. No. 1.2e-77;
  Matches  177;  Conservative   41;  Mismatches   88;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||:||||| ||||||  ||:|||||| :::||:|::||  || | |:||:  |    :|
Db        560 VFLFPGQGSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRV 619

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||||||:|||||| ||||||:|| ||:|||||||||||| |||:|:||||:|  |||
Db        620 DVVQPVLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRL 679

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              : :: | ||||:| |     | : :   |  || ||||||||| |||:  |:     ||
Db        680 MRSLSGEGGMAAVALGEAAVRERL-RPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCA 738

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
               |:| | | |||||| | :: :|||||  |||:|:|: | | ||||   :| | |||
Db        739 AEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTELDAR 798

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTE--DTAEDVTAIGSLR 298
              |||||| : |||:|||  ||: |:  |:|||||| :| |:|:  |  | |||   :|||
Db        799 YWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAEDAVVVGSLH 858

Qy        299 RGDNDTRRF 307
```

```
Db        859 RDGGDLSAF 867


RESULT 10
AAY92707
ID   AAY92707 standard; Protein; 4150 AA.
XX
AC   AAY92707;
XX
DT   29-AUG-2000  (first entry)
XX
DE   S. antibioticus 8,8a-deoxyoleandolide synthase I encoded by oleAI.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   Region          9..1034
FT                   /label= Loading_module
FT   Domain          9..428
FT                   /label= KS-Q_domain
FT                   /note= "inactivated ketosynthase domain"
FT   Domain          562..931
FT                   /label= AT_domain
FT                   /note= "malonyl-specific acyl transferase domain"
FT   Domain          954..1034
FT                   /label= ACP_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          1061..2526
FT                   /label= Extender_module_1
FT   Domain          1061..1478
FT                   /label= KS1_domain
FT                   /note= "ketosynthase domain"
FT   Domain          1580..1926
FT                   /label= AT1_domain
FT                   /note= "acyl transferase domain"
FT   Domain          2162..2349
FT                   /label= KR1_domain
FT                   /note= "ketoreductase domain"
FT   Domain          2440..2526
FT                   /label= ACP1_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          2545..4069
FT                   /label= Extender_module_2
FT   Domain          2545..2973
FT                   /label= KS2_domain
FT                   /note= "ketosynthase domain"
FT   Domain          3075..3420
FT                   /label= AT2_domain
FT                   /note= "acyl transferase domain"
FT   Domain          3700..3884
FT                   /label= KR2_domain
FT                   /note= "ketoreductase domain"
FT   Domain          3979..4075
FT                   /label= ACP2_domain
FT                   /note= "acyl carrier protein domain"
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24478.
XX
PR   29-OCT-1998;   98US-0106100.
PR   16-FEB-1999;   99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   N-PSDB; AAA09469.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 27-28; 86pp; English.
XX
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence   4150 AA;

  Query Match             53.9%;  Score 861.5;  DB 21;  Length 4150;
  Best Local Similarity   56.7%;  Pred. No. 2.8e-75;
  Matches  174;  Conservative   35;  Mismatches   95;  Indels    3;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||   || : ||||  :  |   |     ||:|::||   |:||  |
Db       3076 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDRLTGWSLVEVLR--GGEAVLGRV 3133

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| |::|||||  || ||:|| ||:||||||||| | | ||| | |: | |||||
Db       3134 DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRA 3193

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              :| : | ||| || | |: |  :: : : ||: ||||||| || |||||:|:||:|| |
Db       3194 IARIAGGGGMVSVGLSAERVRTML-DTYGGRVSVAAVNGPSSTVVSGDAQALDELLAGCE 3252

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
               ||||||:|||||||   :  ||:|||| | |::|| | |||||||   ||||| |||
Db       3253 REGVRARRVPVDYASHSAQMDQLRDELLEALADVTPQDSSVPFFSTVTADWLDTTALDAG 3312

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||: || : ||| :||: |    |   ||| |||| ||| | :| :    |  |: ||||
Db       3313 YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGITETLDTFDADAVALSSLRRD 3372

Qy        301 DNDTRRF 307
              :    ||
Db       3373 EGGLDRF 3379


RESULT 11
```

```
AAY92709
ID   AAY92709 standard; Protein; 3519 AA.
XX
AC   AAY92709;
XX
DT   29-AUG-2000  (first entry)
XX
DE   S. antibioticus 8,8a-deoxyoleandolide synthase III encoded by oleAIII.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   Region          33..1564
FT                   /label= Extender_module_5
FT   Domain          33..466
FT                   /label= KS5_domain
FT                   /note= "ketosynthase domain"
FT   Domain          569..915
FT                   /label= AT5_domain
FT                   /note= "acyl transferase domain"
FT   Domain          1199..1383
FT                   /label= KR5_domain
FT                   /note= "ketoreductase domain"
FT   Domain          1478..1564
FT                   /label= ACP5_domain
FT                   /note= "acyl carrier protein domain"
FT   Region          1686..3218
FT                   /label= Extender_module_6
FT   Domain          1686..2115
FT                   /label= KS6_domain
FT                   /note= "ketosynthase domain"
FT   Domain          2220..2566
FT                   /label= AT6_domain
FT                   /note= "acyl transferase domain"
FT   Domain          2855..3039
FT                   /label= KR6_domain
FT                   /note= "ketoreductase domain"
FT   Domain          3132..3218
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein domain"
FT   Domain          3290..3519
FT                   /label= TE_domain
FT                   /note= "thiol esterase domain"
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24478.
XX
PR   29-OCT-1998;   98US-0106100.
PR   16-FEB-1999;   99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   N-PSDB; AAA09469.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 29; 86pp; English.
XX
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence   3519 AA;

  Query Match             53.7%;  Score 858;  DB 21;  Length 3519;
  Best Local Similarity   57.0%;  Pred. No. 4.9e-75;
  Matches  175;  Conservative   34;  Mismatches   94;  Indels    4;  Gaps    3;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR-RDAGDAVWER 59
              |||||||| || ||   || : ||||  :  |   | |   ||:|::|: |||   | |
Db       2221 VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDA--TVLGR 2278

Qy         60 ADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 119
               ||||| |::|||||  || ||:|| ||:|||||||||| | | ||| | |: | ||||
Db       2279 VDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSR 2338

Qy        120 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 179
              |:| : | ||| || ||| |  :: : : ||: ||||||| || |||| :|:||:|| |
Db       2339 AIARIAGGGGMVSVSLPAGRVRTML-DTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGC 2397

Qy        180 AGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDA 239
              ||||||:|||||||   :  ||:|||| | ||:|| | |||||||   ||||| |||
Db       2398 EREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADWLDTTALDA 2457

Qy        240 AYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRR 299
              ||: || : ||| :||: |   |   ||| |||| ||| || | :   ::   :|||||
Db       2458 GYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLRR 2517

Qy        300 GDNDTRR 306
              :   |
Db       2518 DEGGLDR 2524


RESULT 12
AAY39297
ID   AAY39297 standard; Protein; 2595 AA.
XX
AC   AAY39297;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnA a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
```

```
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1944
FT                   /note= "Encoded by CAG"
FT   Domain          6..423
FT                   /label= KSi
FT                   /note= "Beta-ketosynthase domain: part of the
FT                         initiator module"
FT   Domain          528..853
FT                   /label= ATi
FT                   /note= "Acyl transferase domain: part of the
FT                         initiator module"
FT   Domain          895..977
FT                   /label= ACPi
FT                   /note= "Acyl carrier protein domain: part of the
FT                         initiator module"
FT   Domain          998..1413
FT                   /label= KS1
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 1"
FT   Domain          1525..1858
FT                   /label= AT1
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 1"
FT   Domain          2158..2337
FT                   /label= KR1
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 1"
FT   Domain          2432..2513
FT                   /label= ACP1
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 1"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 87-96; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnA gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnA gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   2595 AA;

  Query Match             53.3%;  Score 852;  DB 20;  Length 2595;
  Best Local Similarity   55.2%;  Pred. No. 1.2e-74;
  Matches  169;  Conservative   47;  Mismatches   88;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              ||||||| || ||| :|| :| |||  |  | :|||| :|||::|:||  || |  :
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| ||::||||| ||||:|: | ||:||||||||||| | ||||::|||: ||:|||
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              | |: | | |||:  ||:|| |::   |  |: || |||| || |||| ||: |:|| ||
Db        647 LTALAGSGAMASLQHPAEEVRQIL-LPWRDRIGVAGVNGPSSTLVSGDREAMAELLAECA 705

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :| |||||:|||| ||:: :|:||| ||  : |:   :| :||  |  ||   :||
Db        706 DRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPMDAD 764

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| ||| |  ||:||   |:  |:|:|||| |   |::|      :| |:|:||||
Db        765 YWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAGREVVALGTLRRG 824

Qy        301 DNDTRR 306
              :  |:
Db        825 EGGMRQ 830


RESULT 13
AAB70965
ID   AAB70965 standard; Protein; 2595 AA.
XX
AC   AAB70965;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF18, SEQ ID 41.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
```

```
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88335.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 57; Page 179-189; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   2595 AA;

  Query Match             53.3%;  Score 852;  DB 22;  Length 2595;
  Best Local Similarity   55.2%;  Pred. No. 1.2e-74;
  Matches  169;  Conservative   47;  Mismatches   88;  Indels    2;  Gaps    2;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| || ||| :|| :| |||  |  | :|||| :|||::|:||  || |  :
Db        527 VFVFPGQGSQWIGMGRQLLETSEVFAASMSDCADALAPHLDWSLLDVLRNAAGAAHLDHD 586

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||||| ||::|||| |||||:|: | ||:|||||||||| | ||||::|||: ||:|||
Db        587 DVVQPALFAIMVSLAELWRSWGVRPVAVVGHSQGEIAAACVAGALSVRDAARVVAVRSRL 646

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              | |: | | |||:  ||:|| |::   |  |: || |||| || |||| ||: |:|| ||
Db        647 LTALAGSGAMASLQHPAEEVRQIL-LPWRDRIGVAGVNGPSSTLVSGDREAMAELLAECA 705

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                 :| |||||:|||| ||:: :|:||| ||  : |:   :| :||  |  ||   :||
Db        706 DRELRMRRIPVEYASHSPHIEVVRDELLGLLAPVEPRTGSIPIYSTTTGDLLD-RPMDAD 764

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              |||||| ||| |  ||:||   |:  |:|:|||| |   |::|      :| |:|:||||
Db        765 YWYRNLRQPVLFEAAVEALLKRGYDAFIEISPHPVLTANIQETAVRAGREVVALGTLRRG 824

Qy        301 DNDTRR 306
              :  |:
Db        825 EGGMRQ 830


RESULT 14
AAY39299
ID   AAY39299 standard; Protein; 3170 AA.
XX
AC   AAY39299;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnC a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..423
FT                   /label= KS3
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 3"
FT   Domain          531..850
FT                   /label= AT3
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 3"
FT   Domain          1159..1337
FT                   /label= KR3
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 3"
FT   Domain          1425..1506
FT                   /label= ACP3
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 3"
FT   Domain          1529..1952
FT                   /label= KS4
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                         module 4"
FT   Domain          2066..2396
FT                   /label= AT4
FT                   /note= "Acyl transferase domain: part of extender
FT                         module 4"
FT   Domain          2700..2880
FT                   /label= KR4
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                         module 4"
FT   Domain          2972..3053
FT                   /label= ACP4
FT                   /note= "Acyl carrier protein domain: part of extender
FT                         module 4"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 103-113; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnC gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnC gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
```

```
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   3170 AA;

  Query Match             53.1%;  Score 848.5;  DB 20;  Length 3170;
  Best Local Similarity   56.2%;  Pred. No. 3.7e-74;
  Matches  172;  Conservative   35;  Mismatches   98;  Indels    1;  Gaps    1;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| ||||| | || :|| ||| |  |  ||  :||||:||:||        :|
Db        563 VFVFPGQGWQWAGMAVDLLDASPTFARHMDECATALRRYVDWSLVDVLRGAENSPPLDRV 622

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:|||||| :|||||: | ||:||||||||||   | | |:|||: |||||||
Db        623 DVLQPASFAVMVSLAEVWRSYGVRPAAVVGHSQGEIAAACAAGVLPLEDAARLVALRSRA 682

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |  : |||||||:  || ||  |  .  ||| |||:|||||  |||| |||||:|| ||
Db        683 LKGLSGRGGMASLACPADEVAALFAGS-GGRLEVAAINGPRSVVVSGDLEAVDELLAECA 741

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :||||||||||||  ||: :|  :|     :  :   ||::||| | |:|   ||
Db        742 EKDMRARRIPVDYASHSAHVEVVRSPVLAAAAGVRHRDGQVPWWSTVIGDWVDPARLDGE 801

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||  |||||  |||  | :  |  :|:|:|  || |  |:|:|  ::    | |:|:|||
Db        802 YWYRNLRQPVRFEHAVQGLVERGFGLFIEMSAHPVLTTAVEETGAESETAVAAVGTLRRD 861

Qy        301 DNDTRR 306
                ||
Db        862 SGGLRR 867


RESULT 15
AAB70967
ID   AAB70967 standard; Protein; 3170 AA.
XX
AC   AAB70967;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF20, SEQ ID 46.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88337.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 57; Page 226-238; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   3170 AA;

  Query Match             53.1%;  Score 848.5;  DB 22;  Length 3170;
  Best Local Similarity   56.2%;  Pred. No. 3.7e-74;
  Matches  172;  Conservative   35;  Mismatches   98;  Indels    1;  Gaps    1;

Qy          1 VFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERA 60
              |||||||| ||||| | || :|| ||| |  |  ||  :||||:||:||        :|
Db        563 VFVFPGQGWQWAGMAVDLLDASPTFARHMDECATALRRYVDWSLVDVLRGAENSPPLDRV 622

Qy         61 DVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA 120
              ||:||  |:|||||| :|||||: | ||:||||||||||   | | |:|||: |||||||
Db        623 DVLQPASFAVMVSLAEVWRSYGVRPAAVVGHSQGEIAAACAAGVLPLEDAARLVALRSRA 682

Qy        121 LAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCA 180
              |  : |||||||:  || ||  |       ||| |||:|||||  |||| |||||:|| ||
Db        683 LKGLSGRGGMASLACPADEVAALFAGS-GGRLEVAAINGPRSVVVSGDLEAVDELLAECA 741

Qy        181 GTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA 240
                :||||||||||||  ||: :|  :|     :  :   ||::||| | |:|   ||
Db        742 EKDMRARRIPVDYASHSAHVEVVRSPVLAAAAGVRHRDGQVPWWSTVIGDWVDPARLDGE 801

Qy        241 YWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRG 300
              ||||||  |||||  |||  | :  |  :|:|:|  || |  |:|:|  ::    | |:|:|||
Db        802 YWYRNLRQPVRFEHAVQGLVERGFGLFIEMSAHPVLTTAVEETGAESETAVAAVGTLRRD 861

Qy        301 DNDTRR 306
                ||
Db        862 SGGLRR 867


Search completed: June 17, 2003, 13:02:29
Job time : 16.0391 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 183.577 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3
Perfect score: 20829
Sequence: 1 VQRMDGGEEPRPAAGEVLGV.....FTSATEAEIFKFIDNDLGLS 3972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	20826	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	17634.5	84.7	3626	2	Q9F7T9	Q9f7t9 streptomyce
3	12972	62.3	5532	2	Q9S0R4	Q9s0r4 streptomyce
4	11330.5	54.4	6239	2	Q9S0R7	Q9s0r7 streptomyce
5	8148	39.1	4613	2	Q9ZGI5	Q9zgi5 streptomyce
6	7891	37.9	6048	2	Q93H87	Q93h87 streptomyce
7	7705.5	37.0	4472	2	O33954	O33954 streptomyce
8	7516.5	36.1	6146	2	Q93HJ5	Q93hj5 streptomyce
9	7356	35.3	3613	2	Q93HJ1	Q93hj1 streptomyce
10	7281.5	35.0	5588	2	Q9ALM2	Q9alm2 saccharopol
11	7244	34.8	4340	2	O30764	O30764 streptomyce
12	7221.5	34.7	6797	2	Q9X993	Q9x993 streptomyce
13	7217	34.6	4840	2	Q93HJ4	Q93hj4 streptomyce
14	7190.5	34.5	4881	2	Q9S0R3	Q9s0r3 streptomyce
15	7168	34.4	10917	2	Q93NW6	Q93nw6 streptomyce
16	7098	34.1	5069	2	O52789	O52789 amycolatops
17	7094	34.1	5060	2	O52545	O52545 amycolatops
18	7089.5	34.0	11096	2	Q9L4W3	Q9l4w3 streptomyce
19	7075	34.0	4498	2	Q93HE5	Q93he5 streptomyce
20	7049.5	33.8	4290	2	Q9WXC0	Q9wxc0 micromonosp
21	7041.5	33.8	3816	2	Q9KIV3	Q9kiv3 streptomyce
22	7017	33.7	3739	2	Q9ZGI4	Q9zgi4 streptomyce
23	6986	33.5	3939	2	Q93HJ3	Q93hj3 streptomyce
24	6952.5	33.4	3576	16	Q9EX53	Q9ex53 streptomyce
25	6943.5	33.3	4685	2	Q93HJ2	Q93hj2 streptomyce
26	6939.5	33.3	5435	2	Q9L4X2	Q9l4x2 streptomyce
27	6912.5	33.2	4150	2	Q9KIV4	Q9kiv4 streptomyce
28	6828	32.8	6145	2	Q93H84	Q93h84 streptomyce
29	6813.5	32.7	4151	16	O53490	O53490 mycobacteri
30	6804.5	32.7	3170	2	Q9ALM4	Q9alm4 saccharopol
31	6778.5	32.5	3524	2	Q93H86	Q93h86 streptomyce
32	6756	32.4	10223	2	Q54296	Q54296 streptomyce
33	6753	32.4	9507	2	Q9EWA1	Q9ewa1 streptomyce
34	6728	32.3	8817	2	Q53840	Q53840 polyangium
35	6716.5	32.2	7525	2	Q9KIE0	Q9kie0 streptomyce
36	6652	31.9	9477	2	Q9L4X3	Q9l4x3 streptomyce
37	6642.5	31.9	8563	2	Q54297	Q54297 streptomyce
38	6624.5	31.8	3970	2	Q93HI8	Q93hi8 streptomyce
39	6595.5	31.7	9510	2	Q93NX9	Q93nx9 streptomyce
40	6579.5	31.6	3654	2	O30766	O30766 streptomyce
41	6555	31.5	3729	2	O33956	O33956 streptomyce
42	6469.5	31.1	5644	2	Q93NX8	Q93nx8 streptomyce
43	6447	31.0	4928	2	Q9ALM3	Q9alm3 saccharopol
44	6416	30.8	7576	2	Q9ZGA4	Q9zga4 streptomyce
45	6302.5	30.3	3546	2	Q9F830	Q9f830 micromonosp

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 20826; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3971; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRAC 60
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRAC 60

Qy         61 EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHS 120

Qy        121 LGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNP 180

Qy        181 ARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVS 240

Qy        241 GLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVE 300

Qy        301 MSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATT 360

Qy        361 VLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVE 420

Qy        421 LRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD 480

Qy        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540

Qy        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600

Qy        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660

Qy        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720

Qy        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780

Qy        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGG 840

Qy        841 VIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 VIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGI 900

Qy        901 GGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960

Qy        961 VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREEL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREEL 1020

Qy       1021 VQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 VQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQA 1080
```

```
Qy       1081 CEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGH 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGH 1140

Qy       1141 SQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 SQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGR 1200

Qy       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260

Qy       1261 GDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320

Qy       1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380

Qy       1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVES 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVES 1440

Qy       1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL 1500

Qy       1501 PTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 PTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHT 1560

Qy       1561 RQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 RQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATT 1620

Qy       1621 NATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPH 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 NATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPH 1680

Qy       1681 HQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 HQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPT 1740

Qy       1741 QHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVI 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVI 1800

Qy       1801 HTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQ 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 HTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQ 1860

Qy       1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920

Qy       1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLA 1980

Qy       1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040

Qy       2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100

Qy       2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160

Qy       2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220
```

```
Db       2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220
Qy       2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280
Db       2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280
Qy       2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340
Db       2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340
Qy       2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400
Db       2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400
Qy       2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460
Db       2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460
Qy       2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520
Db       2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520
Qy       2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580
Db       2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580
Qy       2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640
Db       2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640
Qy       2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700
Db       2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700
Qy       2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760
Db       2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760
Qy       2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQ 2820
Db       2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQ 2820
Qy       2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880
Db       2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880
Qy       2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940
Db       2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940
Qy       2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000
Db       2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000
Qy       3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060
Db       3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060
Qy       3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120
Db       3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120
Qy       3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180
Db       3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180
Qy       3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
Db       3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
Qy       3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
Db       3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
Qy       3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360
Db       3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360
Qy       3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420
Db       3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420
Qy       3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480
Db       3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480
Qy       3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540
Db       3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540
Qy       3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600
Db       3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600
Qy       3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660
Db       3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660
Qy       3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720
Db       3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720
Qy       3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780
Db       3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780
Qy       3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840
Db       3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840
Qy       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLPQPDNAVAP 3900
Db       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLPQPDNAVAP 3900
Qy       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960
Db       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960
Qy       3961 IFKFIDNDLGLS 3972
Db       3961 IFKFIDNDLGLS 3972
```

RESULT 2
Q9F7T9
ID Q9F7T9 PRELIMINARY; PRT; 3626 AA.
AC Q9F7T9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Avermectin polyketide synthase (Fragment).
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31271;
RA Hong Y.-S., Lee J.J.;
RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT and polyketide synthase gene from Streptomyces avermitilis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275943; AAG09812.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.

```
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT   NON_TER    3626   3626
SQ   SEQUENCE   3626 AA;  380557 MW;  6272F5F088C1A8D0 CRC64;

  Query Match             84.7%;  Score 17634.5;  DB 2;  Length 3626;
  Best Local Similarity   92.6%;  Pred. No. 0;
  Matches 3419;  Conservative   33;  Mismatches  115;  Indels  125;  Gaps   21;

Qy         20 VADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDS  79
              :|||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db          1 MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAEFAPYVDWSVEQVLRDS  60

Qy         80 PDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAA 139
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAA 120

Qy        140 RVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGA 199
              ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGA 180

Qy        200 REAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVT 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 REAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVT 240

Qy        260 GGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLG 319
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        241 GGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAADLG 300

Qy        320 DTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLL 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLL 360

Qy        380 DLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIF 439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIL 420

Qy        440 DHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSP 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSP 480

Qy        500 KDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGI 559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 KDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGI 540

Qy        560 SPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEG 619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEG 600

Qy        620 AADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLA 679
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |
Db        601 AADVEGHILTGTTGSVLSGRSAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFA 660

Qy        680 LAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLG 739
              |||||| :  ||||||||| |||||||||| ||||||||||:|:
Db        661 LAGGVSDVHP-GMFIEFSRQCGLSVDGRCKAVSAAADGTGWGEGIGL------------ 706

Qy        740 HRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLG 799
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 ---LAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLG 763

Qy        800 DPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHV 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        764 DPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHV 823

Qy        860 DKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCV 919
              |:||||   :||||||||||||||||||||||||||||||||||||||||||||||||||

Db        824 DEPSPQ-GLAAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCV 882

Qy        920 AGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEA 979
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 AGGRVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEA 942

Qy        980 LRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAA---------------------- 1015
              |||| ||:|||||||||||||| :||||  |:||| .
Db        943 LRQYEAVQPDVSLADVGAGLACRQAVLEHHVVILAACTSSSRSAARTTARSSSTARPQAR 1002

Qy       1016 DREELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFA 1075
                    : :||||||     :||||||||||||||||| ||||||||||||    |||
Db       1003 PAPTSARRIGALAAGSGSAALTTGHAPGGDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFA 1062

Qy       1076 RRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPD 1135
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||:
Db       1063 RRMQACEEALAPWVDWSVVDILRRDAGDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPN 1122

Qy       1136 AVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGE 1195
              |||||: ||||||: ||||||||||||                  || |||||||||
Db       1123 EVLGHSKDEIAAAHIYGALSLKDAAKTVA------------------LPPQEVEQLIGE 1163

Qy       1196 RWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREE 1255
              |  |||||||||||||||||||||| ||||||||||  ||||||||||||||||||||
Db       1164 R-GGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRP-RIPVDYASHCPHVQPLREE 1221

Qy       1256 LLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRV 1315
              |||||||||||| ||||||||||||||||||||||||||||||||  |||||| ||||
Db       1222 LLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSHDVQALADQGHRV 1281

Qy       1316 FVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWH 1375
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1282 LLEVSPHPTLVPAIEDTTEDTAR---------------RR--------HCDRQPPPRRER 1318

Qy       1376 H-----------HYTH---HHTHP--HPHTHLDLPT-------YPFQHQHYWLESSQPGA 1412
              |           :| |   ||  |   |  ||  |          ||||||||||||||||
Db       1319 HPLLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA 1378

Qy       1413 GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPA 1438

Qy       1473 LSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLH 1532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 LSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLH 1498

Qy       1533 HHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHT 1592
              ||||||||| |  ||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 HHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHT 1558

Qy       1593 PTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPT 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 PTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPT 1618

Qy       1653 HTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPT 1712
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1619 HTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPT 1678

Qy       1713 PHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTI 1772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 PHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTI 1738

Qy       1773 TTCDTSNPDQL-QQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH 1831
              |||||| |       |||||||||:|||||| ||||||||||||||||||||||||||||
Db       1739 TTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPTQLNNVLRAKAHSAH 1798

Qy       1832 LLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGT 1891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 LLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGT 1858

Qy       1892 WQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQ 1951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 WQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQ 1918
```

```
Qy      1952 TSDLVSAARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAI 2011
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1919 TSDLVSAARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAI 1978

Qy      2012 APDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAES 2071
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1979 APDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAES 2038

Qy      2072 APSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDT 2131
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2039 APSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDT 2098

Qy      2132 LYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAG 2191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2099 LYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAG 2158

Qy      2192 INPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPA 2251
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2159 INPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPA 2218

Qy      2252 VSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKA 2311
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2219 VSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKA 2278

Qy      2312 FSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQR 2371
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2279 FSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQR 2338

Qy      2372 VIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNV 2431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2339 VIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNV 2398

Qy      2432 GHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRL 2491
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2399 GHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRL 2458

Qy      2492 RRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKS 2551
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2459 RRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKS 2518

Qy      2552 QPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAA 2611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2519 QPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAA 2578

Qy      2612 GEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTH 2671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2579 GEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTH 2638

Qy      2672 LDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYA 2731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2639 LDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYA 2698

Qy      2732 GHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHEND 2791
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2699 GHSLGEITGAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHEND 2758

Qy      2792 LAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHT 2851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2759 LAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHT 2818

Qy      2852 QTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDN 2911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2819 QTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDN 2878

Qy      2912 TLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHTHTH- 2970
             |||||||||||| |||||||||||||||||||||||||||||||||||  ||  | |
Db      2879 TLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHH-AQPTPHPHP 2937

Qy      2971 LDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRS 3030
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2938 LDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRS 2997

Qy      3031 HPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVA 3090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2998 HPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVA 3057

Qy      3091 AADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGA 3150
             ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3058 AADGKGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGA 3117

Qy      3151 VAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALL 3210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3118 VAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALL 3177

Qy      3211 DGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGRE 3270
             :|||||||||||||||:|||||:|||||||||||||||||||||||||||||||||||||||
Db      3178 NGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGRE 3237

Qy      3271 AVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQA 3330
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3238 AVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQA 3297

Qy      3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3298 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3357

Qy      3391 AGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTG 3450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3358 AGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTG 3417

Qy      3451 AGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVP 3510
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      3418 AGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRSGDTRARRSDVP 3477

Qy      3511 AQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSR 3570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3478 AQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSR 3537

Qy      3571 RGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDD 3630
             |||||||||||      |||||||||||||||||||  |    | ::  |      |
Db      3538 RGPDAPGAEGLRRSWPR-GAEVRIVACDVGERREVVRLAGG--CSCRVSVFV------DF 3588

Qy      3631 ATIASLTPERLGTVFAAKVDAAL---LLDELT 3659
             : :||:|  ||      : :||:    ::| |
Db      3589 SAVASVT-VRLPVASDVRKEAAMAYATVEEFT 3619
```

RESULT 3
Q9S0R4
ID Q9S0R4 PRELIMINARY; PRT; 5532 AA.
AC Q9S0R4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 3.
GN AVEA3.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84478.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.

```
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5532 AA;  575193 MW;  EC7B2069D44A4057 CRC64;

  Query Match             62.3%;  Score 12972;  DB 2;  Length 5532;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2640;  Conservative  168;  Mismatches  519;  Indels  366;  Gaps   47;

Qy       478 AADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGT 537
Db        35 ASSEPIAIVGMACRFPGGVCSPEELWELVASGGDAIGEFPAGRGWDLEGLFDSDPDRSGT  94

Qy       538 FYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRT 597
Db        95 SYARYGGFLYEAGEFDADFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSMRGSRT 154

Qy       598 GVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGC 657
Db       155 GVFAGVMYHDYGSRLGTI-PEG---FEGYIGNGSGGAVASGRVAYTLGLEGPAVSVDTAC 210

Qy       658 SASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADG 717
Db       211 SSSLVALHLACQSLRSGECTLALAGGVTVMSTPHLFVEFSRQRGLSVDGRCKSFAGGADG 270

Qy       718 TGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALA 777
Db       271 TGMGEGVGMLLVERLSDAVRLGHRVLAVLRGSAVNQDGASNGLTAPNGPAQERVIRQALA 330

Qy       778 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAG 837
Db       331 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGNRPLWLGSVKSNIGHAQAAAG 390

Qy       838 VGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 897
Db       391 VGGVIKMVMALREGVLPRTLHVDEPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 450

Qy       898 FGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 957
Db       451 FGVSGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 510

Qy       958 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 1017
Db       511 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 570

Qy      1018 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 1077
Db       571 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 630

Qy      1078 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAV 1137
Db       631 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAV 690

Qy      1138 LGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERW 1197
Db       691 LGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERW 750

Qy      1198 AGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELL 1257
Db       751 AGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELL 810

Qy      1258 ELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFV 1317
Db       811 ELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFV 870

Qy      1318 EVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHH 1377
Db       871 EVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHH 930

Qy      1378 YTHHHTHPHPHTHLDLPTYPFQHQHYWLES---------------------------- 1407
```

```
Db       931 YTHHHTHPHNH-HLDLPTYPFQRQHYWLDAPTGAGDVAAAGLEPAEHPLLAATVQLADTD 989

Qy      1408 ----------SQPGAG----------SGS---------------------------- 1416
Db       990 GCLLTGRLSLRSHPWLGDYEVGGAVLLSGSAFVELAVQVGERVGCTRIEQLTVHAPLVVP 1049

Qy      1417 ----------------------------GAGAGSGAGSGRAGT-----------AGGTAE 1437
Db      1050 VGGGVSVQVGVAAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGGVV 1109

Qy      1438 VE--SRFW---DAVARQDLETVATTLAVPPSAG--LDTVVPALSAWHRHQHDQARINTWT 1490
Db      1110 VDGLAGVWPPRGAVA-VDVDGVRDRLA---GAGCVLGPVFSGLRAVWRDGGD-LLAEVCL 1164

Qy      1491 YQETW--------KPLTLPTTHQPHQTWL---------------IAIPE----TQTHHPH 1523
Db      1165 PEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAG 1224

Qy      1524 ITNILTNLHHHG----------------------------ITPIPL--------------- 1541
Db      1225 VTGVRVRVWAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGS 1284

Qy      1542 --TLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALD------------ETPHP 1587
Db      1285 LYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGVGEV---- 1341

Qy      1588 HHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTL 1647
Db      1342 ------VGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGQEDGPVDVVGAAVWGLVRSAQ 1395

Qy      1648 LEHPTHTAGIIDLPT-TPTPHTLQ----------------HLTQTLT--QPHHQTQLAIR 1688
Db      1396 AEHPDRFV-LLDLDTDTDTGTDLDTGAGAGAGAGWGVDGGHVAAVVACGEP----QLAVR 1450

Qy      1689 TTGTHTRRLTPTTLTPTHQPPTP-------------TPHGTTLITGGTGALATHLTHHLT 1735
Db      1451 GERVLAARL--TRLESSVDVPAQRSGDVAGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 1508

Qy      1736 THQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHP 1795
Db      1509 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 1568

Qy      1796 LTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATF 1855
Db      1569 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 1628

Qy      1856 GAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNG----LADSDKARAYLDRR 1910
Db      1629 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRR--IIRS 1686

Qy      1911 GFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPT 1970
Db      1687 GLHPMSTPDALALFDAALA-LDRPVLLPADL--------RPAPPLPPLLQDLLPATRRRT 1737

Qy      1971 P----------PAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRAL 2020
Db      1738 TRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDL 1797

Qy      2021 GFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG---TAAESAPSAAA 2077
Db      1798 GFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLLGSDSTASIPAPRAAA 1857

Qy      2078 VTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db      1858 VPADQD--EPVAIIGMACRYPGGVTSAEELWELLASGRDTVGEFPTDRGWDLEALFDPEP 1915

Qy      2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db      1916 GRPGTSYTRCGSFLYDAGEFDAGFFGISPREALAMDPQQRLLLEASWEAMEQAGIDPTTV 1975

Qy      2198 HGTPTGVFTGTNGQDYALRVH-NAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDT 2256
```

```
Db       1976 RGSQTGVFAGLIPQAYGPRLHENAAADTEGYVLTGTSGSVASGRISYTFGFEGPAVSVDT 2035

Qy       2257 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2316
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2036 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2095

Qy       2317 DGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2376
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       2096 DGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2155

Qy       2377 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2156 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2215

Qy       2437 AAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2496
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db       2216 AAGVAGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2275

Qy       2497 SSFGVSGTNAHVILEEA---------PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLV 2547
              |||||||||||||||||         |||||||||||||| |||| :|||||||||||||||
Db       2276 SSFGVSGTNAHVILEEAPAHNIPSDTPADDVPGGPPAGE-DAGSGEEAAAGSPGVWPWLV 2334

Qy       2548 SAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ 2607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 SAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ 2394

Qy       2608 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2454

Qy       2668 ICTHLDPHLDHPLLPLLTQNDNDN-----EDAAALLQQTRYAQPALFAFQVALHRLLTDG 2722
              |||||||||||||||: |       |:|||||||| ||||||||||||||||||||||||||
Db       2455 ICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQPALFAFQVALHRLLTDG 2514

Qy       2723 YHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHIT 2782
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2515 YHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHIT 2574

Qy       2783 HHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNP 2842
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2575 HHITAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNP 2634

Qy       2843 ILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVT 2902
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       2635 ILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDIATTTQTLHQHGVT 2694

Qy       2903 TYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHD 2962
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       2695 TYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHH 2754

Qy       2963 NQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALL 3022
              |||||||||||||||||||||||||||||||||||||||||||||||||||    | ||
Db       2755 NQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELAEGDGCLL 2814

Qy       3023 AGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGG 3082
              ||||||:||||| ||||| ||| |  | ||||||| |||| |||||||| |||:|  |
Db       2815 TGRLSLRTHPWLAGHAVGGVVLLPGTAFAELALHAGESVGCDHVDELTLHTPLVIPEVGD 2874

Qy       3083 VSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVV-VDGL 3141
              |::|| :|| |  |||:::::::||     ||  ||    || ||:|||  |         : |
Db       2875 VTLQVAIAAPDESGRRMMTIHSRGEGGSGGADASASAWTRHAAGVLSPAKDDDTASYELL 2934

Qy       3142 AGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAA 3201
              || ||| ||  ||::   |::| ||   |  | |||| || | |:| |  ||||  ||||
Db       2935 AGPWPPVGATPVDLNTAYDQMADAGFAYGLAFQGLRAAWRYGDDILVEARLPEEVSGDAA 2994

Qy       3202 GFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVS 3261
              :|||||||  :|       |||    || |  | || || |:     |   ||  :|
Db       2995 AYGLHPALLDAALQ--------GTGLLSVAGPGTPV-VPHVWNGLRFRTHGAVSVRACLS 3045

Qy       3262 AVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV 3321
              :|  |   || | : |: |||||||||||||||||||||||||||||||||||||||||||
Db       3046 TLGATGA--AVCVRITDDTGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV 3103

Qy       3322 GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVV 3381
              |||||||||||||||||||||||||||||||||||||||||||      ||||||||||||||||
Db       3104 GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGA----DGGGGGGVGEVVGGVLGVV 3159

Qy       3382 QGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT 3441
              |||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db       3160 QGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGAAVWGLVRSAQAEHPDRFVLLDLDT 3219

Qy       3442 --DTGTDLDTG--AGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQ 3497
                :| | | |  || | |:|||||||||||||||||||||||:|||||  ||||
Db       3220 DLDSGADADAGNEAGMGSGLDGGRVAAVVACGEPQLAVRGERVLAARLTRLES------- 3272

Qy       3498 RSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 3557
                                        ||||||||||||||||||||||||||||||||
Db       3273 ----------------------PVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 3308

Qy       3558 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3309 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3368

Qy       3618 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 3677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3369 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 3428

Qy       3678 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGL 3737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3429 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGL 3488

Qy       3738 HPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADN 3797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3489 HPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADN 3548

Qy       3798 GAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELR 3857
              |||||||||||||||||||||||||||||||||||||||||||||  |: || | ||||||||||||||
Db       3549 GAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHNAPEMIPVDSAFRDLGFDSLTAVELR 3608

Qy       3858 NRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL 3890
              |||   |||||||:| || ||  ||  ||  :|
Db       3609 NRLGEATGLRLPTSLVFDQPNAATLARHLRREL 3641
```

```
RESULT 4
Q9S0R7
ID   Q9S0R7      PRELIMINARY;      PRT;  6239 AA.
AC   Q9S0R7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 2.
GN   AVEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84475.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
```

```
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6239 AA;  666283 MW;  A706280940B502FA CRC64;

  Query Match             54.4%;  Score 11330.5;  DB 2;  Length 6239;
  Best Local Similarity   44.3%;  Pred. No. 0;
  Matches 2512;  Conservative  312;  Mismatches  961;  Indels 1883;  Gaps   61;

Qy          1 VQRMDGGEEPRPA-------AGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVF 53
Db        565 LQALAAG-EPHPAVIHSSAPGGTGTG---EAAGKTAFICSGQGTQRPGMAHGLYHTHPVF 620

Qy         54 RESVRACEAAFAPYVDWSVEQVLRDSPD----------APGLDRVDVVQPTLFAVMISL- 102
Db        621 AAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALH 680

Qy        103 AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVA 162
Db        681 RLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMP-PGTMTTLH 739

Qy        163 ATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPV 222
Db        740 TTPHHITHHLTAHEND----LAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPT 795

Qy        223 DVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVR 282
Db        796 NHAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQLLTPHYWTQQARNTVD 852

Qy        283 FEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDH 342
Db        853 YATTTQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLT-------LTHPHHHPQTH 905

Qy        343 FLTSLAQ-----------------------------------LRGHGE----------- 355
Db        906 LLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGRE 965

Qy        356 --TSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGF 413
Db        966 QATTPSTPLRDVLVGKSPQERDEELLRLVRTHAAAVL----GHATPEVIVPNKAFKELGF 1021

Qy        414 DSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGH 473
Db       1022 DSLAAIQLRNRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLT---- 1077

Qy        474 GGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPT 533
Db       1078 --AVPADEPIAIVGMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPD 1135

Qy        534 QPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVR 593
Db       1136 HHGTSYTRSGGFLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLR 1195

Qy        594 GSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITV 653
Db       1196 GTPTGVFAGLTYHDYAARFPTA----PAGFEGYLGHGSAGSIASGRVAYALGLEGPALTV 1251

Qy        654 DTGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSA 713
Db       1252 DTACSSSLVALHLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSA 1311

Qy        714 AADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIR 773
Db       1312 AADGTGWGEGVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIR 1371

Qy        774 QALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHT 832
Db       1372 LALANADLTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHA 1431

Qy        833 MAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRR 892
Db       1432 QAAAGVGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPG-GEGRLRR 1490

Qy        893 AGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVA-------- 944
Db       1491 AGVSSFGVSGTNAHVILEEAPAHN--IPSDTPADDAPGEA--AADDVPGEAAGDDAGTGG 1546

Qy        945 -VSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGR 1003
Db       1547 EATGPAAGS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHAR 1601

Qy       1004 AVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGHAPGGD-----RGGVVFVFPGQGG 1058
Db       1602 AVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGT 1661

Qy       1059 QWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGD----------AVWER 1108
Db       1662 QRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQ 1721

Qy       1109 ADVVQPVLFSVMVSL-AALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRS 1167
Db       1722 TPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 1781

Qy       1168 RALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAY 1227
Db       1782 -TLMQTMPPGTMTTLHTTPHHITHHLTAH-ENDLAIAAINTPTSLVISGTPHTVQHITTL 1839

Qy       1228 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 1287
Db       1840 CQQQGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQLLT 1896

Qy       1288 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 1347
Db       1897 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNT--PTTTLTLTH 1954

Qy       1348 RGDNDTRRFLTALAHTHTTGIGTPTTWH-HHYTHHHTHPHPHTHLDLPTYPFQHQHYWLE 1406
Db       1955 PHHHPQTHLLTNLAK-------TTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLE 2007

Qy       1407 SSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGL 1466
Db       2008 STQPGAGS------GSGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGL 2061

Qy       1467 DTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITN 1526
Db       2062 DTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITN 2121

Qy       1527 ILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH 1586
Db       2122 ILTNLHHHGITPIPLTLNHTHTNPQH----LHHTRQQAQNHTTGPITGLLSLLALDETPH 2177

Qy       1587 PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTT 1646
Db       2178 PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTT 2237

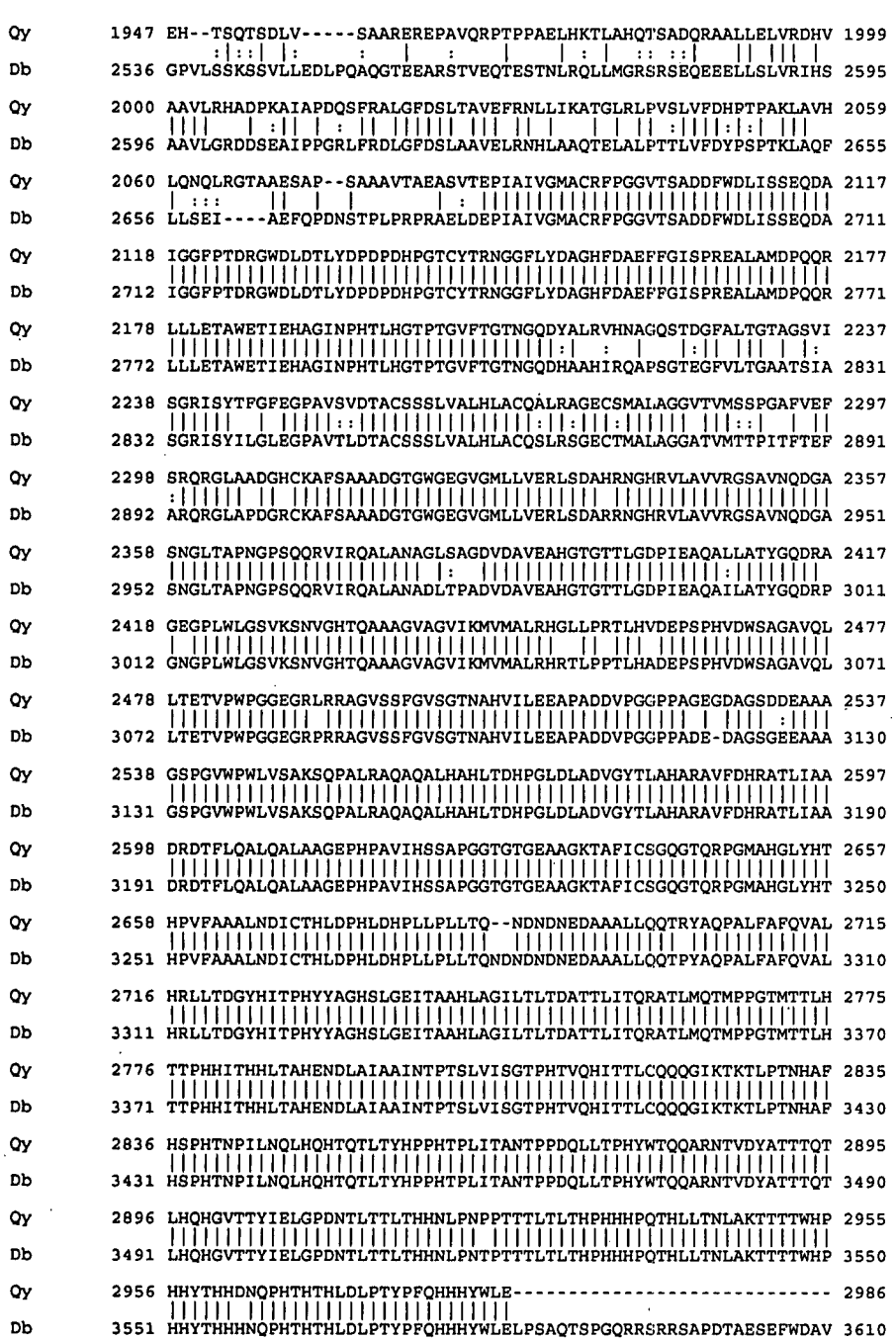
Qy       1647 LLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTH 1706
Db       2238 LLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTH 2297

Qy       1707 QPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQK 1766
Db       2298 QPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQK 2357

Qy       1767 GIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAK 1826
Db       2358 GIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAK 2417

Qy       1827 AHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATS 1886
Db       2418 AHSAHLLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATS 2477

Qy       1887 IAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKI 1946
Db       2478 IAWGTWQGNGLA-TGQVSEHLRRRGMFAMPPELAVTAVDGAIA-SGRPSLLVADIDWKKL 2535
```


Qy 1947 EH--TSQTSDLV-----SAAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHV 1999
Db 2536 GPVLSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHS 2595
Qy 2000 AAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVH 2059
Db 2596 AAVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQF 2655
Qy 2060 LQNQLRGTAAESAP--SAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDA 2117
Db 2656 LLSEI----AEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDA 2711
Qy 2118 IGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQR 2177
Db 2712 IGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQR 2771
Qy 2178 LLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVI 2237
Db 2772 LLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAATSIA 2831
Qy 2238 SGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEF 2297
Db 2832 SGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPITFTEF 2891
Qy 2298 SRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGA 2357
Db 2892 ARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGA 2951
Qy 2358 SNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRA 2417
Db 2952 SNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAILATYGQDRP 3011
Qy 2418 GEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQL 2477
Db 3012 GNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSPHVDWSAGAVQL 3071
Qy 2478 LTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAA 2537
Db 3072 LTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPADE-DAGSGEEAAA 3130
Qy 2538 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 2597
Db 3131 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 3190
Qy 2598 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 2657
Db 3191 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 3250
Qy 2658 HPVFAAALNDICTHLDPHLDHPLLPLLTQ--NDNDNEDAAALLQQTRYAQPALFAFQVAL 2715
Db 3251 HPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAALLQQTPYAQPALFAFQVAL 3310
Qy 2716 HRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLH 2775
Db 3311 HRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLH 3370
Qy 2776 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 2835
Db 3371 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 3430
Qy 2836 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQT 2895
Db 3431 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQT 3490
Qy 2896 LHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHP 2955
Db 3491 LHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHP 3550
Qy 2956 HHYTHHDNQPHTHTHLDLPTYPFQHHHYWLE---------- 2986
Db 3551 HHYTHHHNQPHTHTHLDLPTYPFQHHHYWLELPSAQTSPGQRRSRRSAPDTAESEFWDAV 3610


Qy 2987 ---------- 2986
Db 3611 NEEDLQSLAETLDIDASALDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPH 3670
Qy 2987 ---------- 2986
Db 3671 QTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNH 3730
Qy 2987 ---------- 2986
Db 3731 TTGPITGLLSLLALDETPHPHHPHTPTGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHP 3790
Qy 2987 ----------STQP-------- 2990
Db 3791 NDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAI 3850
Qy 2991 ---------- 2990
Db 3851 RTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 3910
Qy 2991 ---------- 2990
Db 3911 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNL 3970
Qy 2991 ----------GAGN------- 2994
Db 3971 FAPVSETDAESFSSVTAAKATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAA 4030
Qy 2995 ---------- 2994
Db 4031 NAYLDALATHRQTHGLPGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHR 4090
Qy 2995 ---------- 2994
Db 4091 ARAQRPNSNLIIADIDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQP 4150
Qy 2995 ---------- 2994
Db 4151 IATSLRERLARLTSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQF 4210
Qy 2995 ---------- 2994
Db 4211 SKELAKETGLPLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVED 4270
Qy 2995 ---------- 2994
Db 4271 EPIAIIGMACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYT 4330
Qy 2995 ---------- 2994
Db 4331 RNGGFLYGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 4390
Qy 2995 ---------- 2994
Db 4391 AGINAQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLV 4450
Qy 2995 ---------- 2994
Db 4451 ALHWAAQALRAGECSMALAGGVTVMSSPGTFVEFSRQRGLAADGRCKAYSAAADGTGWAE 4510
Qy 2995 ---------- 2994
Db 4511 GVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLT 4570
Qy 2995 ---------- 2994
Db 4571 PADVDAVEGHGTGTTLGDPIEAQALLAAYGQHRPHHRPLWLGSLKSNIGHAQAAAGVGGV 4630
Qy 2995 ---------- 2994
Db 4631 IKMVMALRNGLLPQTLHVDEPTPQVDWSTGAVQLLTQPVPWPADPAGRPRHAGVSSFGVS 4690
Qy 2995 ---------- 2994

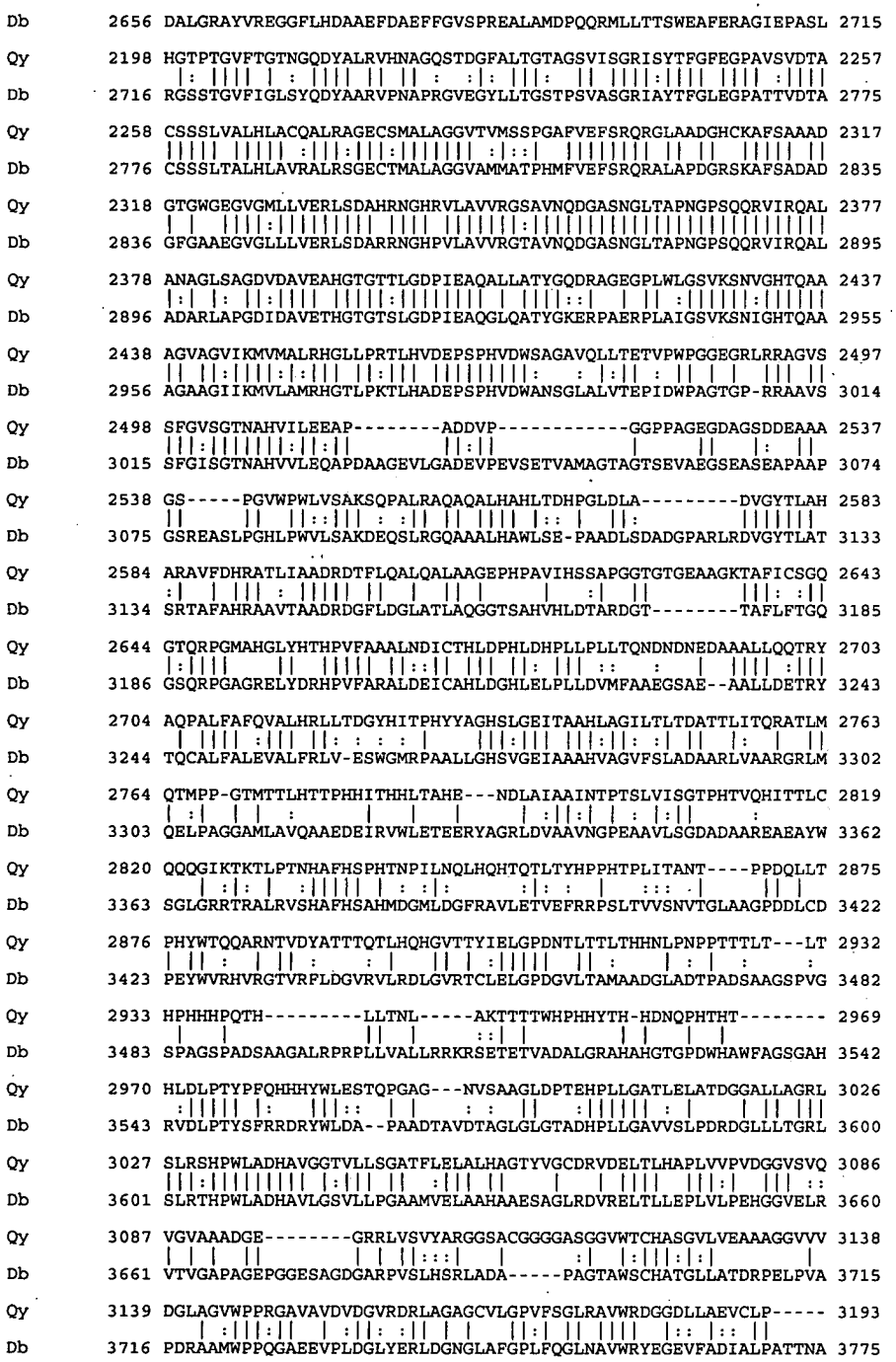
```
Db       4691 GTNAHIILEEAPTPQDSDTDDEPPANAPALPHPLPLPVPVSARSEAGLRAQAQALRQYVA 4750
Qy       2995 ------------------------------------------------------------ 2994
Db       4751 ARPDMSPADIGAGLARGRAVLEHRAVILAADREELAQALTALAAGEPHPHITTGHTRGGD 4810
Qy       2995 ------------------------------------------------------------ 2994
Db       4811 RGGVVFVFPGQGGQWAGMGLTLLTSSPVFAEHIDACEKALTPWVPWSLTDILHRDPDDPA 4870
Qy       2995 ------------------------------------------------------------ 2994
Db       4871 WQQADVVQPVLFSIMVSLAALWRSYGIEPDAVLGHSQGEIAAAHICGALSLKDAAKTVAL 4930
Qy       2995 ------------------------------------------------------------ 2994
Db       4931 RSRALAAVRGRGAMASLPLPAQDVQQLISERWEGQLWVAALNGPHSTTVSGDTKAVDEVL 4990
Qy       2995 ------------------------------------------------------------ 2994
Db       4991 AHCTDTGLRAKRIPVDYASHCPHVQPLHDELLHLLGDITPQPSTVPFFSTVEGTWLDTTT 5050
Qy       2995 ------------------------------------------------------------ 2994
Db       5051 LDAAYWYRNLHQPVRFSHAIQTLTDDGHRAFIEISPHPTLVPAIEDTTENTTENITATGS 5110
Qy       2995 ------------------------------------------------------------ 2994
Db       5111 LRRGDNDTHRFLTALAHTHTTGIGTPTTWHHHYTQTHPHPNPHTHLDLPTYPFQHQHYWL 5170
Qy       2995 --------VSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLS 3046
Db       5171 QPPTTTTDLTTTGLTPTHHPLLTATLTLADNNTQLLTGRLSLRTHPWLTDHTVAGMVLLP 5230
Qy       3047 GATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARG 3106
Db       5231 GTALLELALQAGERVDCPRVEELTLHAPLVIPHTEDVTLQVTVRAADESGHRALAIHSYS 5290
Qy       3107 GSACGGGGASGGVWTCHASGVLVEAAAGGVVVDG-----LAGVWPPRGAVAVDVDGVRDR 3161
Db       5291 GTA----SSADREWTRHATGLLTHHADTDHRADTHTDACLGGSWPPPGAQPIELGDVYGR 5346
Qy       3162 LAG-AGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGD--AAGFGLHPALLDGVVQPLS 3218
Db       5347 MAADSDIAYGPVFQGLHAAWRFGDDVLAEVRLPEEALRDAPAAAFGVHPALLDAALHATA 5406
Qy       3219 VLLPGGTGFGEGAG-------FGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREA 3271
Db       5407 LTPQNGDGSTENVAQESMPDRAAHQARLPFSWSGVSLHTAGSSVLRVRLSRSPQHG--NA 5464
Qy       3272 VSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV-----GPVPV 3326
Db       5465 VALTAADEDGRPVVTIESLALRPVSTEELRAAADRTPEHESLFRLDWVSVPVPANAPSPT 5524
Qy       3327 CGQAWA------------WHEDV-----------GESGGGPVPGVVVL--------RCP 3354
Db       5525 ADRPWAVIGAGLPHLPGLTEHEHVTAYDEPADLLLALDRGAPPPGVLVVGGVAHTEAREY 5584
Qy       3355 DAGAGGGGGGGGGGGVGEVV-------------GGVLGVVQGWLGLERFAGSRLVVVTR 3400
Db       5585 SAEAPGERGTEACEARPDVVHVGVVHTAAVHAAAAQMLARLQAWLGDERLADSRLLVLTC 5644
Qy       3401 GAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGG 3460
Db       5645 GAVARASGDDATDLPGAAVWGLVRSAQSEHPDRITLLDF--ERGTE----------AEPG 5692
Qy       3461 RVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPA- 3519
Db       5693 QLATALNCGERQLAVRPGGLFTPRLVR--------APRVADA------VPAVPAVAVPSA 5738
Qy       3520 -RRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA 3578
Db       5739 GHAAVPAAG----PFLPGGTVLITGGTGVLGRLVARHLVEAHGVRHLLLAGRRGPDAEGA 5794
Qy       3579 EGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTP 3638
Db       5795 PELRAELGGLGATVEVVACDAADRQQLADLLTRIPDDRPLTGVVHSAGILDDGVITSLSP 5854
Qy       3639 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 3698
Db       5855 ERLGAVLRAKADAALLLDELTRGAELSAFVMFSSASAVVGSPGQGNYAAANAVLDFLAHR 5914
Qy       3699 RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRP 3758
Db       5915 RRAEGLPAVSLAWGLWEEGTGMTGHLDVDDHARISRAGMRPLPTAEALALFDAALADGEP 5974
Qy       3759 VLLPADL--------RPAPPLPPLLQDLLPATRRRTTRTTTTGG---ADNGAQLHARLAG 3807
Db       5975 FLMPARLDLTAVRSGAASAPVPPLLQGLLQLPRSRSAAAAPGHGAPAADEAAAWRERLAR 6034
Qy       3808 QTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLR 3867
Db       6035 QSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAATGLR 6094
Qy       3868 LPTTLAFDHPNPTTLTHHLHTQLQPQPD 3895
Db       6095 LRATLAFDFPTPAALAEHLGERLLPDQE 6122
```

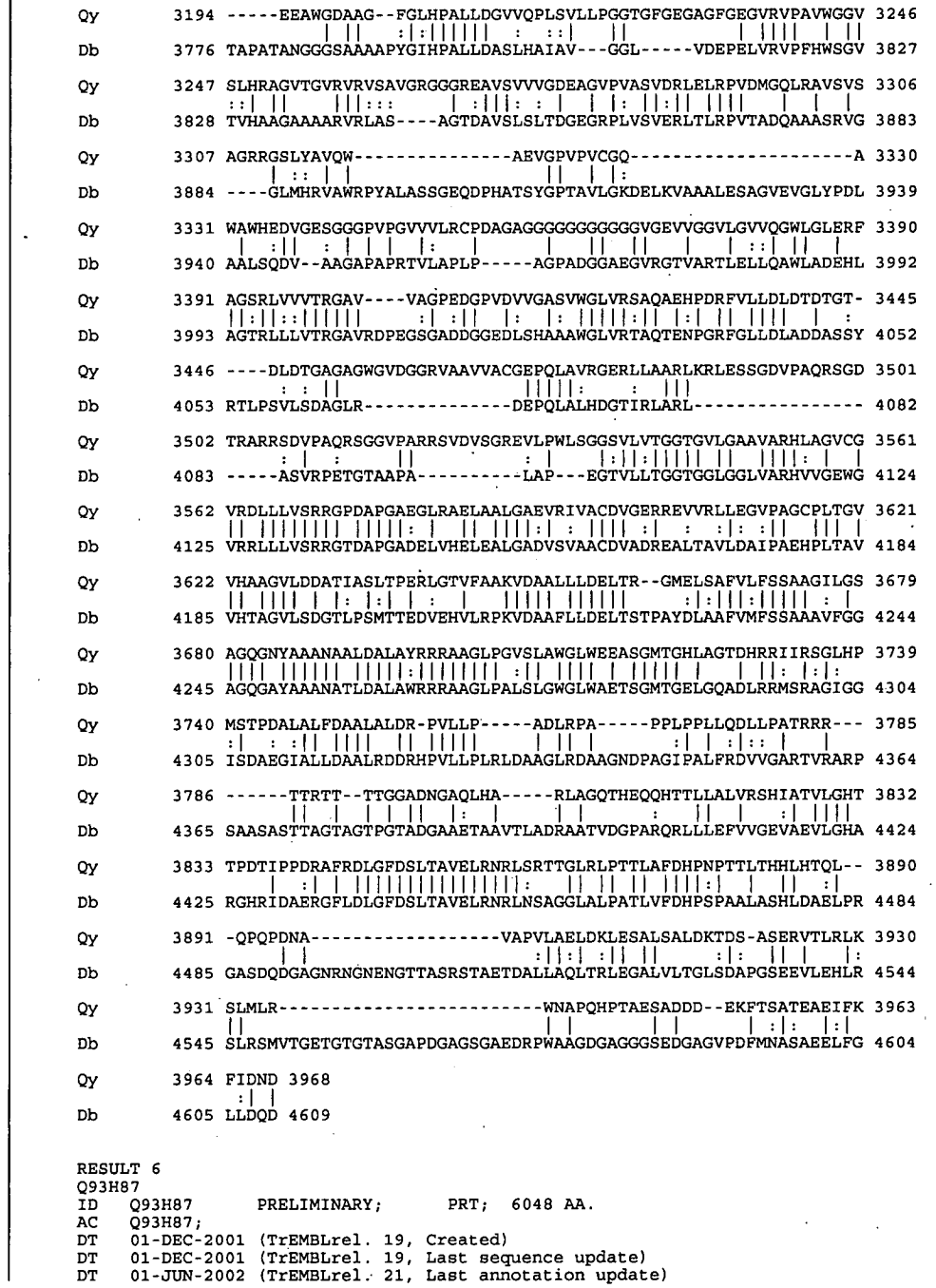
RESULT 5
Q9ZGI5
ID Q9ZGI5 PRELIMINARY; PRT; 4613 AA.
AC Q9ZGI5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase PikAI.
GN PIKAI.
OS Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=54571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC15439;
RX MEDLINE=98445333; PubMed=9770448;
RA Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT venezuelae: architecture of metabolic diversity.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL; AF079138; AAC69329.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 4613 AA; 477281 MW; A3BAF8D37CEC9383 CRC64;

 Query Match 39.1%; Score 8148; DB 2; Length 4613;
 Best Local Similarity 44.8%; Pred. No. 0;
 Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

```
Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
Db        620 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 678
```







RESULT 6
Q93H87
ID Q93H87 PRELIMINARY; PRT; 6048 AA.
AC Q93H87;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```
DE   Modular polyketide synthase.
GN   PTEA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69303.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
KW   Phosphopantetheine.
SQ   SEQUENCE   6048 AA;  629420 MW;  F8673C2ED2694705 CRC64;

  Query Match             37.9%;  Score 7891;  DB 2;  Length 6048;
  Best Local Similarity   45.7%;  Pred. No. 0;
  Matches 1838;  Conservative  443;  Mismatches 1269;  Indels  474;  Gaps   83;

Qy         26 GGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVL---RDSPDA 82
              |  |:| ||| |  |||||| ::   |  ::    |   | :   : :|:     : :|
Db        548 GELAFLFSGQGSQRAGMGRELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEA 607

Qy         83 PGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVV 142
                ||:    | :|||| ::|  |  |  : |  ::|||:||:|||||:| ||||||  :|
Db        608 AELDQTLYTQTSLFAVEVALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALV 667

Qy        143 TLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREA 202
                 :    |  ||:::| |  ||:  ||   |:    |:::||:||| | ||||  :
Db        668 AARGRLMQALPAGGAMIAVQAAEDEVRARIGERTD----RVSIAALNGPDSVVVSGDEDL 723

Qy        203 VADLVADLTAAQVRTRMIPVDVPAHSP---LMYAIEERVVSGLLPITPRPSRIPFHSSVT 259
               ::      |   :|  : |    |||    | |     |  ||   |    |||  |:||
Db        724 ATEIADAFAALGRKTSRLRVSHAFHSPHMEPMLAEFRNVAEGL---TFHAPRIPIVSNVT 780

Qy        260 GGRLDTRE-----LDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQEL 314
               |||           |||| |::   |||       | :|| :|::|: |  ||
Db        781 -GRLSEEPSGYEGCTAAYWVRHVREAVRFADGVARLDEQGVRTYLELGPGGVLA------ 833

Qy        315 APDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHG-------------------- 354
                 :      :::: :  ||  :       | ::| |  ||
Db        834 --SMARAGADSESLFVPALRTRRPEAQALLAAVASLHVHGVEPDWDALFGGGRGARRKVA 891

Qy        355 ------------------------------------ETSATTV--LSARLTALSPTQQQ 375
                                                   || :  |  |  ||  |: :::   .
Db        892 LPTYAFERQRYWLDGDTTSPVATTGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERD 951

Qy        376 SLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPV 435
               :||:|||| |||        |       :|  |||||:  |||||:||:|||||||
Db        952 EAVLDMVRAHIAAVL----GHAEARQVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPS 1007

Qy        436 TLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAA---------DDPIAII 486
               |:||| ||||:|  |   :||                || |||           :||||:
Db       1008 GLLFDHPTPAALARHLAAQSLG----------------GGDAAAAAPATVADPHEPIAIV 1051

Qy        487 GMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFL 546
                :|| |||| ||:||| | ||||| |  |||||||  |  :  :|  ||    : ||||
Db       1052 AMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWDVETLYDPEPGTPGKSSTRHGGFL 1111

Qy        547 HDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSF 606
               |:|| ||||||||||||| |:||||||:|||:|||||||||| ::|||| ||| || :
Db       1112 HEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEAFERAGIDPATLRGSRAGVFIGATAQ 1171

Qy        607 DYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHL 666
               |||||:     | |  |||::||||| || |||:||||||||||:|||| ||:||| |||
Db       1172 DYGPRL----HEPADGVEGYLLTGTTTSVASGRVAYSFGLEGPAVTVDTACSSSLVALHL 1227

Qy        667 ACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGM 726
               | |||||||||:||||||:|||| |||:|||||||| ||||||::| |||||| |||||
Db       1228 AVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRGLSADGRCKAFAAGADGTGWAEGVGM 1287

Qy        727 LLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADV 786
               |||||||| | ||:|||||||||||||||||||||||:|:||:|||| |||: |||
Db       1288 LLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPSQQRVVRQALAAAGLAPADV 1347

Qy        787 DVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMV 845
               | ||  |||| |||||||:||:| ||| | ||||||||||||||||| |||| ||||||
Db       1348 DAVEAHGTGTLLGDPIEAEALIAAYGQGRDGDRPLWLGSLKSNIGHTQAAAGVAGVIKMV 1407

Qy        846 MALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNA 905
               ||:| ||||||||||:|:| |||||  |||||||| ||  |    ||| |||||| ||||
Db       1408 MAMRHGVLPRTLHVDEPTPHVDWSASGVRLLTEAVEWP--ARDLPRRAAVSSFGISGTNA 1465

Qy        906 HVILEEA-----PAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960
               |||:|:|     || |  |||                                      |
Db       1466 HVIVEQAATSEEPARGTAVAG------------------------------------P 1487

Qy        961 VPVPVSARSEAGLRAQAEALRQYV--------AVRPDVSLADVGAGLACGRAVLEHRAVV 1012
               ||  |||:|:  || ||  |  ::          |||:    ||   ||  ||||| || |
Db       1488 VPWVVSAKSDVALREQAGRLLSFLDAEGVGGAGVRPE----DVALSLATSRAVLERRAAV 1543

Qy       1013 LAADREELVQGLGALAAGEPDRRVTTGHAP-GGDRGG-VVFVFPGQGGQWAGMGVRLLAS 1070
               :  |  || ||| |||:|       :| || ||  || | |:| ||| |  |||  | |:
Db       1544 VGGDLAELRQGLEALASG------ASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAA 1597

Qy       1071 SPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSY 1130
               ||||    |  : :  ::|  |         :    || ||:|  |:|  |  |:
Db       1598 YPVFA----AAYDEVCAFLDAPV------DVDSEELNQTGTTQPALFAVEVALFRLLESW 1647

Qy       1131 GIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVE 1190
               || || | ||| |||||||| | |||:|||| |: |:  : |:   | | :|    | |
Db       1648 GIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKLVSARAALMQALPDGGAMVAV----QATE 1703

Qy       1191 QLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQ 1250
                  :      : :|||||||  |||  :||  |    || | :  |: | :| | | :
Db       1704 DEVLRHLTDGVGIAAVNGPRSVVVSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMD 1763

Qy       1251 PLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALAD 1310
               |: ||   ::  ::   | :|  | : |   :  | :  || |:: : |||:| :: | :
Db       1764 PMLEEFAAVVRGLTFSESRIPVVSNLTGRLAEVYTPE--YWVRHVREAVRFADGIETLGE 1821

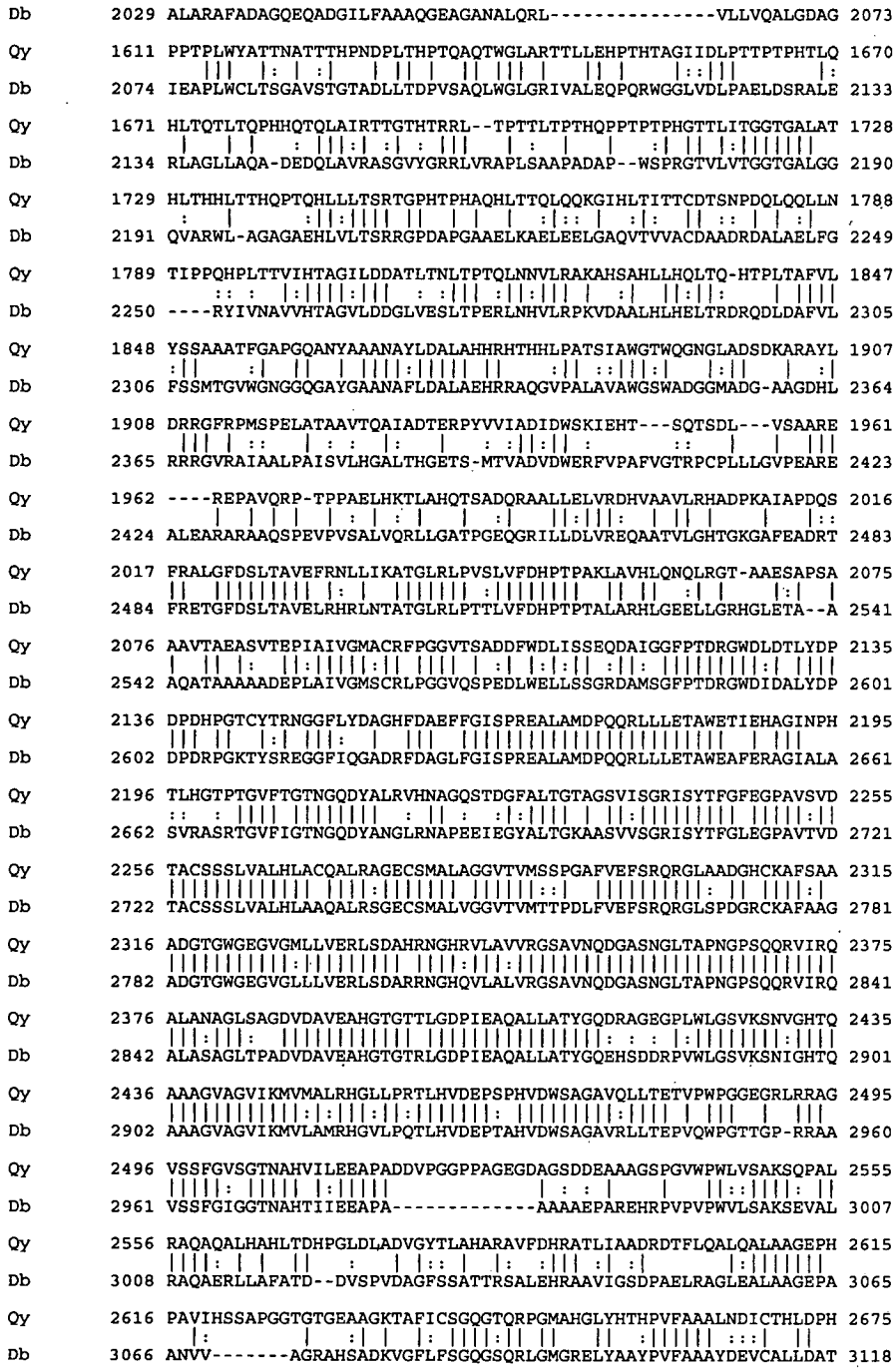
Qy       1311 DGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGT 1370
               |   |||| |   |  |:     |   |:  :  ||    :    :||:   || ||
Db       1822 LGVTTFVEVGPGGVL-SALAQGCFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGIS- 1879

Qy       1371 PTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAG 1430
                  |  :             :||||| || : |||::                   ||
Db       1880 -PDWQVLFPGAR-------RVDLPTYAFQRERYWLDAP------------------RAA 1912

Qy       1431 TAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWT 1490
               |||    ::: |||:|  :|  ::   | : |:  || |    |||| | : :|:  : |
Db       1913 TAGAPDAMDAEFWDSVESEDRASLGALLGLEPTE-LDVVASKLSAWRRKRREQSTTDGWR 1971

Qy       1491 YQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNP 1550
               |: ||:||       |  |||  : |       |   || |   |:  :|  :
Db       1972 YRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTEL---GVALVPFAVGEDTDRD 2028

Qy       1551 QHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH 1610
                     ::||      | |     ||                   : | |
```


Db 2029 ALARAFADAGQEQADGILFAAAQGEAGANALQRL--------------VLLVQALGDAG 2073
Qy 1611 PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQ 1670
Db 2074 IEAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRALE 2133
Qy 1671 HLTQTLTQPHHQTQLAIRTTGTHTRRL--TPTTLTPTHQPPTPTPHGTTLITGGTGALAT 1728
Db 2134 RLAGLLAQA-DEDQLAVRASGVYGRRLVRAPLSAAPADAP--WSPRGTVLVTGGTGALGG 2190
Qy 1729 HLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLN 1788
Db 2191 QVARWL-AGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELFG 2249
Qy 1789 TIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQ-HTPLTAFVL 1847
Db 2250 ----RYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVL 2305
Qy 1848 YSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYL 1907
Db 2306 FSSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADG-AAGDHL 2364
Qy 1908 DRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHT---SQTSDL---VSAARE 1961
Db 2365 RRRGVRAIAALPAISVLHGALTHGETS-MTVADVDWERFVPAFVGTRPCPLLLGVPEARE 2423
Qy 1962 ----REPAVQRP-TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS 2016
Db 2424 ALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEADRT 2483
Qy 2017 FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGT-AAESAPSA 2075
Db 2484 FRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLETA--A 2541
Qy 2076 AAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDP 2135
Db 2542 AQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDIDALYDP 2601
Qy 2136 DPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPH 2195
Db 2602 DPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFERAGIALA 2661
Qy 2196 TLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVD 2255
Db 2662 SVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFGLEGPAVTVD 2721
Qy 2256 TACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAA 2315
Db 2722 TACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLSPDGRCKAFAAG 2781
Qy 2316 ADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQ 2375
Db 2782 ADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLTAPNGPSQQRVIRQ 2841
Qy 2376 ALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQ 2435
Db 2842 ALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDRPVWLGSVKSNIGHTQ 2901
Qy 2436 AAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAG 2495
Db 2902 AAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLLTEPVQWPGTTGP-RRAA 2960
Qy 2496 VSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPAL 2555
Db 2961 VSSFGIGGTNAHTIIEEAPA------------AAAAEPAREHRPVPVPWVLSAKSEVAL 3007
Qy 2556 RAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPH 2615
Db 3008 RAQAERLLAFATD--DVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALAAGEPA 3065
Qy 2616 PAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPH 2675
Db 3066 ANVV-------AGRAHSADKVGFLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT 3118
Qy 2676 LDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSL 2735
Db 3119 VDVD--------------AETLHQTGSTQPALFAVEVALFRLL-ESWGIRPDYVAGHSV 3162
Qy 2736 GEITAAHLAGILTLTDATTLITQRATLMQTMPP-GTMTTLHTTPHHITHHLTAHENDLAI 2794
Db 3163 GEIAAAHVAGVLSLEDAAKLVSARAALMKALPAGGVMVAVQAAEDEMLPHLI---DRAGI 3219
Qy 2795 AAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTL 2854
Db 3220 AAINGPRSVVISGAEDAVTAIAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGL 3279
Qy 2855 TYHPPHTPLITANTPP-DQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTL 2913
Db 3280 TFSEPWIPVVSNLTGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVL 3339
Qy 2914 TTLTHHNLPN--------------PPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYT 2959
Db 3340 SALVQGCLDDGTDIVTIPALRADRPEPQALTTAYAQLH--------VSGAEVDWHAFF-- 3389
Qy 2960 HHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGG 3019
Db 3390 -----PGAR-RVELPTYAFQRERYWLDA-PASAGDMRAVGQGEAGHPLLGAAVPLADSVG 3442
Qy 3020 ALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPV 3079
Db 3443 HLLTGRLSKHTHPWLVDHAVSGVALLPGTAFVELAITAADAVGCDLLEELTLETPLLVPE 3502
Qy 3080 DGGVSVQVGVAAADGEGRRLVSVYAR--GGSACGGGGASGGVWTCHASGVLVEAAAGGVV 3137
Db 3503 RGGVALQVHVGADDGSGRRPLTVHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGG--P 3560
Qy 3138 VDGLAGVWPPRGAVADVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAW 3197
Db 3561 VDGSLTAWPPVGAEPVDVDGFYDRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAER 3620
Qy 3198 GDAAGFGLHPALLD--------GVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLH 3249
Db 3621 TDAGAYGLHPALFDAALHTVWLGAVEP------------EAGTGNGL-LPFAWSGVRLA 3666
Qy 3250 RAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGR 3309
Db 3667 AAGASDLRVKVSRAGT--GTSTVSLVLADGTGGPVAEVDSLTLRPVPADQLRA---GSGS 3721
Qy 3310 RGSLYAVQWAEVG-PVPVCGQAWAWHEDVG----------ESGGGPVPGVVVLRCPDAG 3357
Db 3722 DDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRATDPPADTEADAAAAADAVVVPCPT-- 3779
Qy 3358 AGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGA 3417
Db 3780 ---GTGADLATRVREVTHAVLELVQWWLAEER--PGRLVLVTRGG----------DLAQS 3824
Qy 3418 SVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAV----VACGEPQL 3473
Db 3825 AVCGLVRSAQTENPDRIVLVE------TSSDAHAGAGSGVDDADLARVLPGLIASGEPQA 3878
Qy 3474 AVRG-ERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLP 3532
Db 3879 AVHADDEVRVPRLVRV-------------------------AVP-------DAEPELP 3904
Qy 3533 WLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEV 3592
Db 3905 GLGSGTVLLTGASGGLGGLFARHLVAEHGVRSLLLVSRRGGDAPGAAELTADLTAQGADV 3964
Qy 3593 RIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAA 3652
Db 3965 TWAACDVADRAAVRALLAG--PGQSLSAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAA 4022
Qy 3653 LLLDELTRGM--ELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLA 3710
Db 4023 LNLHDLAGELSGDLSAFVLFSSVAGTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLA 4082

```
Qy       3711 WGLW--EEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADL--- 3765
Db       4083 WGLWAQDSESAMTGSLDHTDLTRIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLG 4142

Qy       3766 ---RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVR 3822
Db       4143 GLQGTVPPVLRGLRGTVRGTARRTAQTAPKG------SLGQRLAGLAEAEREQVVLDLVR 4196

Qy       3823 SHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTL 3882
Db       4197 TEVAAVLGHAGAQSVGADDSFKDIGFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTL 4256

Qy       3883 THHL 3886
Db       4257 AHPL 4260
```

RESULT 7
O33954
ID O33954 PRELIMINARY; PRT; 4472 AA.
AC O33954;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tylactone synthase starter module and modules 1 & 2.
GN TYLG.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RA DeHoff B.S., Sutton K.L., Rosteck P.R. Jr.;
RT "Sequence of Streptomyces fradiae tylactone synthase gene tylG.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U78289; AAB66504.1; -.
DR InterPro; IPR002106; AAtRNA_ligaseII.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR001993; Mitoch_carrier.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 4472 AA; 463275 MW; E911D1B9E0C332CF CRC64;

Query Match 37.0%; Score 7705.5; DB 2; Length 4472;
Best Local Similarity 44.3%; Pred. No. 0;
Matches 1849; Conservative 453; Mismatches 1380; Indels 495; Gaps 97;

```
Qy         13 AAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSV 72
Db        571 AAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVGWDL 630

Qy         73 EQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGG 132
Db        631 LEVV---SGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGA 687

Qy        133 LSLADAARVVTLWS-QAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGP 191
Db        688 LSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKW-----PGVEVAAVNGP 742

Qy        192 RSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIE---ERVVSGLLPITPR 248
```

```
Db        743 ASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSG---IRPR 799

Qy        249 PSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLT 308
Db        800 SPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLV 859

Qy        309 MGLQELAPDLGDTTGTADTVI--MGTLRRGQGTLDHFLTSLAQLRGHGET---------- 356
Db        860 HAIEQ-------TAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPG 912

Qy        357 SATTV-----------------------------LSARLTALSPTQQQ-SLLLDLVRA 384
Db        913 HLTTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR 972

Qy        385 HTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTP 444
Db        973 HTAVVLGLD----TAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTP 1028

Qy        445 AAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAA-----ADDPIAIIGMACRFPGGVRSP 499
Db       1029 RALAQFLHTELL----------DGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSP 1078

Qy        500 KDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGI 559
Db       1079 EALWRLVVEGVDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGI 1138

Qy        560 SPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEG 619
Db       1139 SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRL----AEG 1194

Qy        620 AADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLA 679
Db       1195 AEGSDGYLLTGTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMA 1254

Qy        680 LAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLG 739
Db       1255 LAGGVTVMAGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLG 1314

Qy        740 HRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLG 799
Db       1315 HPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLG 1374

Qy        800 DPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLH 858
Db       1375 DPIEAQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLH 1434

Qy        859 VDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGC 918
Db       1435 VDEPSRHVDWSAGGVWLLTEAREWPG--VDRPRRAAVSAFGVSGTNAHLILE-------- 1484

Qy        919 VAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV--PVPVPVPVPVSARSEAGLRAQ 976
Db       1485 -------------APDTAEAESATTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREA 1531

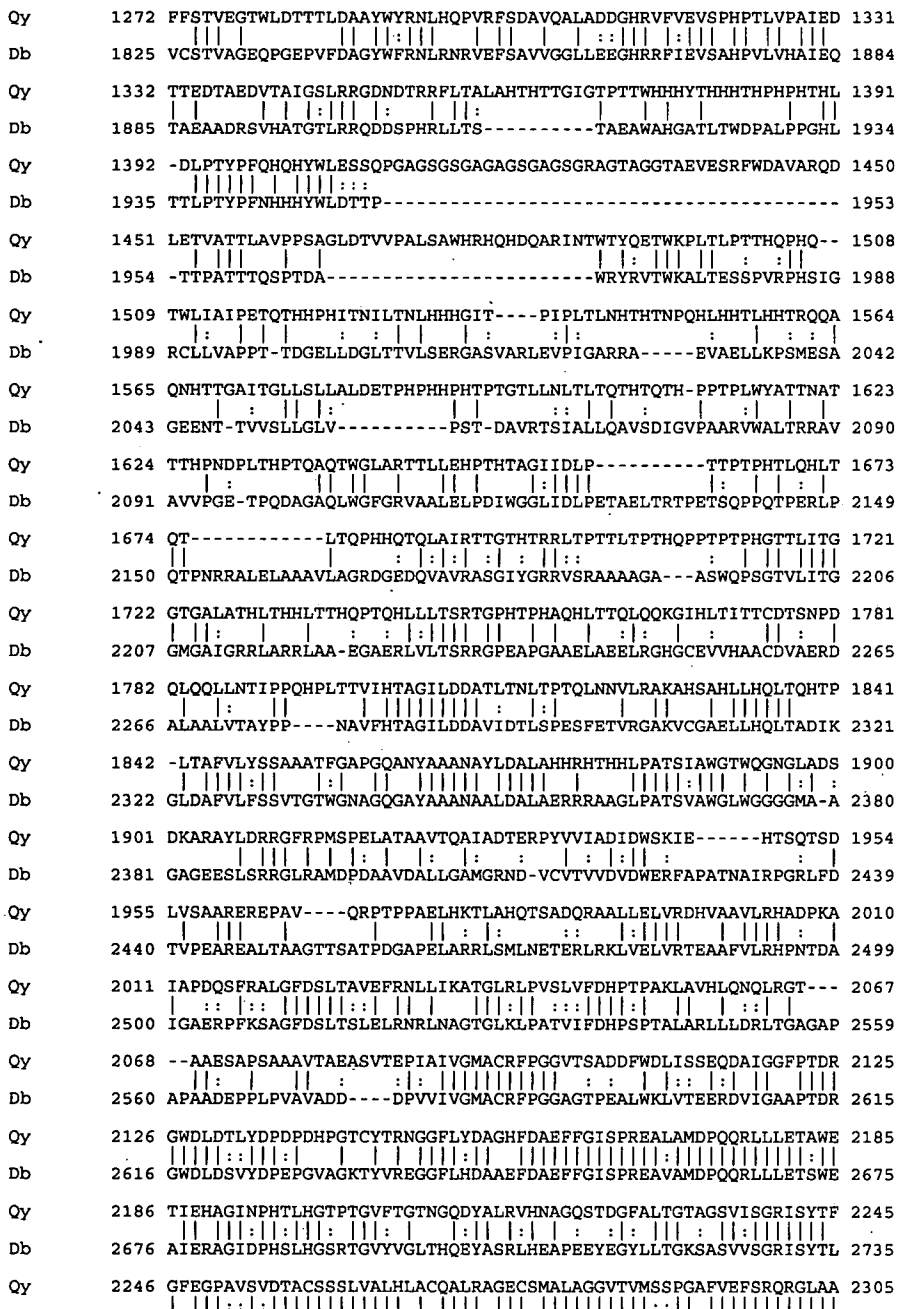
Qy        977 AEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRV 1036
Db       1532 AGRLAEVVEA-GGVGLADVAVTMA-GRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGV 1589

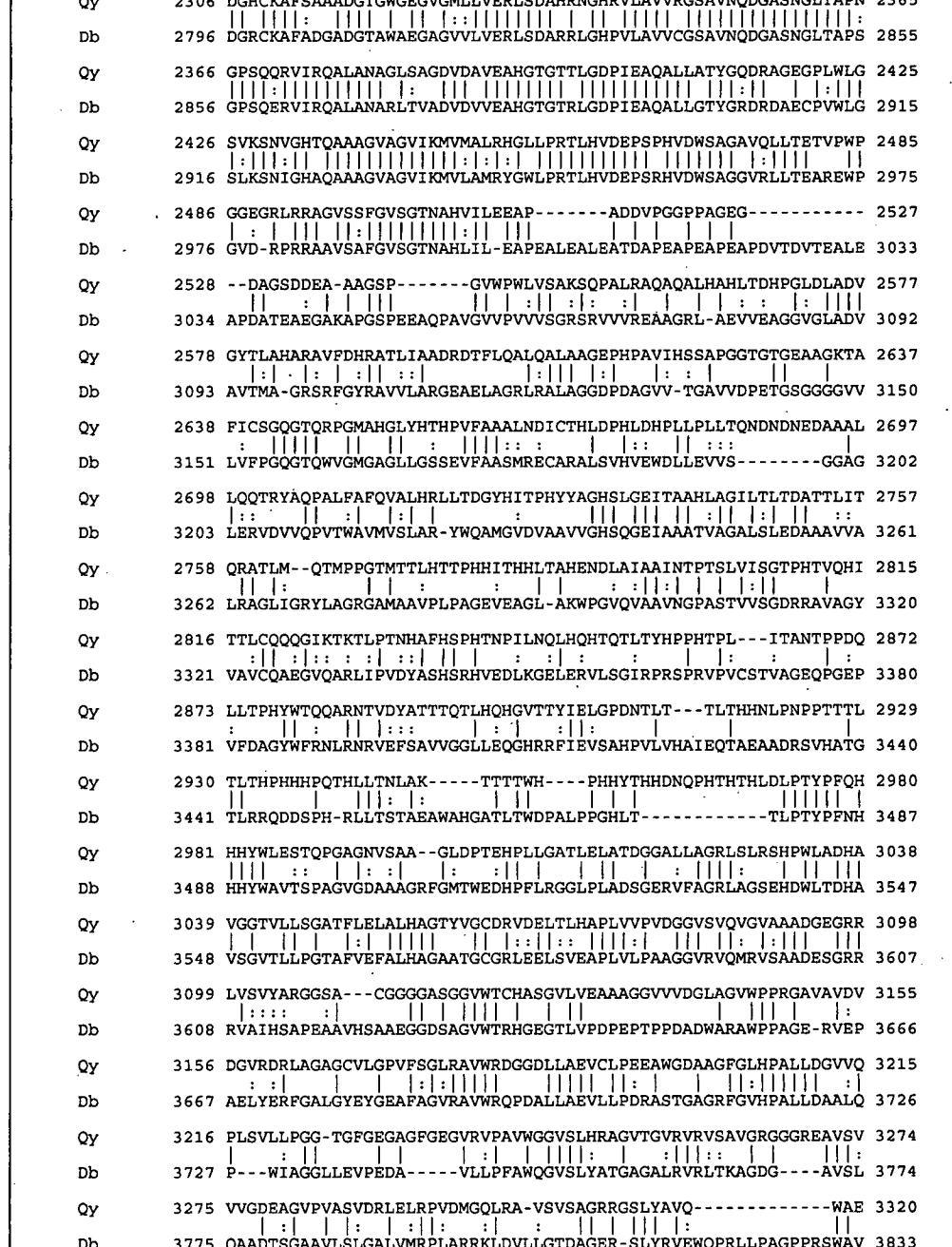
Qy       1037 TTGHA----PGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWS 1092
Db       1590 VTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWD 1649

Qy       1093 VVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCG 1152
Db       1650 LLEVV---SGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAG 1706

Qy       1153 ALSLKDAAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRS 1211
Db       1707 ALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA-KWPG-VEVAAVNGPAS 1764

Qy       1212 TAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVP 1271
Db       1765 TVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVP 1824
```


Qy 1272 FFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIED 1331
Db 1825 VCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQ 1884
Qy 1332 TTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHL 1391
Db 1885 TAEAADRSVHATGTLRRQDDSPHRLLTS----------TAEAWAHGATLTWDPALPPGHL 1934
Qy 1392 -DLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQD 1450
Db 1935 TTLPTYPFNHHHYWLDTTP-- 1953
Qy 1451 LETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQ-- 1508
Db 1954 -TTPATTTQSPTDA----------WRYRVTWKALTESSPVRPHSIG 1988
Qy 1509 TWLIAIPETQTHHPHITNILTNLHHHGIT----PIPLTLNHTHTNPQHLHHTLHHTRQQA 1564
Db 1989 RCLLVAPPT-TDGELLDGLTTVLSERGASVARLEVPIGARRA-----EVAELLKPSMESA 2042
Qy 1565 QNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH-PPTPLWYATTNAT 1623
Db 2043 GEENT-TVVSLLGLV----------PST-DAVRTSIALLQAVSDIGVPAARVWALTRRAV 2090
Qy 1624 TTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLP----------TTPTPHTLQHLT 1673
Db 2091 AVVPGE-TPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAELTRTPETSQPPQTPERLP 2149
Qy 1674 QT-----------LTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITG 1721
Db 2150 QTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAAAGA---ASWQPSGTVLITG 2206
Qy 1722 GTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPD 1781
Db 2207 GMGAIGRRLARRLAA-EGAERLVLTSRRGPEAPGAAELAEELRGHGCEVVHAACDVAERD 2265
Qy 1782 QLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTP 1841
Db 2266 ALAALVTAYPP----NAVFHTAGILDDAVIDTLSPESFETVRGAKVCGAELLHQLTADIK 2321
Qy 1842 -LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS 1900
Db 2322 GLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLPATSVAWGLWGGGGMA-A 2380
Qy 1901 DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIE------HTSQTSD 1954
Db 2381 GAGEESLSRRGLRAMDPDAAVDALLGAMGRND-VCVTVVDVDWERFAPATNAIRPGRLFD 2439
Qy 1955 LVSAARREREPAV----QRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKA 2010
Db 2440 TVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKLVELVRTEAAFVLRHPNTDA 2499
Qy 2011 IAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGT--- 2067
Db 2500 IGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFDHPSPTALARLLLDRLTGAGAP 2559
Qy 2068 --AAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDR 2125
Db 2560 APAADEPPLPVAVADD----DPVVIVGMACRFPGGAGTPEALWKLVTEERDVIGAAPTDR 2615
Qy 2126 GWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWE 2185
Db 2616 GWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGISPREAVAMDPQQRLLLETSWE 2675
Qy 2186 TIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTF 2245
Db 2676 AIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPEEYEGYLLTGKSASVVSGRISYTL 2735
Qy 2246 GFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAA 2305
Db 2736 GLEGPSLSIDTACSSSLVALHNAAQALRGGECDMALAGGVTVMAAPGLFVEFSRQRGLAA 2795


Qy 2306 DGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPN 2365
Db 2796 DGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAVVCGSAVNQDGASNGLTAPS 2855
Qy 2366 GPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLG 2425
Db 2856 GPSQERVIRQALANARLTVADVDVVEAHGTGTRLGDPIEAQALLGTYGRDRDAECPVWLG 2915
Qy 2426 SVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP 2485
Db 2916 SLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVDEPSRHVDWSAGGVRLLTEAREWP 2975
Qy 2486 GGEGRLRRAGVSSFGVSGTNAHVILEEAP-------ADDVPGGPPAGEG----------- 2527
Db 2976 GVD-RPRRAAVSAFGVSGTNAHLIL-EAPEALEALEATDAPEAPEAPEAPDVTDVTEALE 3033
Qy 2528 --DAGSDDEA-AAGSP-------GVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADV 2577
Db 3034 APDATEAEGAKAPGSPEEAQPAVGVVPVVVSGRSRVVVREAAGRL-AEVVEAGGVGLADV 3092
Qy 2578 GYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTA 2637
Db 3093 AVTMA-GRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVV-TGAVVDPETGSGGGGVV 3150
Qy 2638 FICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAAL 2697
Db 3151 LVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVS--------GGAG 3202
Qy 2698 LQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLIT 2757
Db 3203 LERVDVVQPVTWAVMVSLAR-YWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVA 3261
Qy 2758 QRATLM--QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHI 2815
Db 3262 LRAGLIGRYLAGRGAMAAVPLPAGEVEAGL-AKWPGVQVAAVNGPASTVVSGDRRAVAGY 3320
Qy 2816 TTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQ 2872
Db 3321 VAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEP 3380
Qy 2873 LLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLT---TLTHHNLPNPPTTTL 2929
Db 3381 VFDAGYWFRNLRNRVEFSAVVGGLLEQGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATG 3440
Qy 2930 TLTHPHHHPQTHLLTNLAK-----TTTTWH----PHHYTHHDNQPHTHTHLDLPTYPPQH 2980
Db 3441 TLRRQDDSPH-RLLTSTAEAWAHGATLTWDPALPPGHLT------------TLPTYPFNH 3487
Qy 2981 HHYWLESTQPGAGNVSAA--GLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHA 3038
Db 3488 HHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLADSGERVFAGRLAGSEHDWLTDHA 3547
Qy 3039 VGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRR 3098
Db 3548 VSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEAPLVLPAAGGVRVQMRVSAADESGRR 3607
Qy 3099 LVSVYARGGSA---CGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDV 3155
Db 3608 RVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTLVPDPEPTPPDADWARAWPPAGE-RVEP 3666
Qy 3156 DGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQ 3215
Db 3667 AELYERFGALGYEYGEAFAGVRAVWRQPDALLAEVLLPDRASTGAGRFGVHPALLDAALQ 3726
Qy 3216 PLSVLLPGG-TGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSV 3274
Db 3727 P---WIAGGLLEVPEDA-----VLLPFAWQGVSLYATGAGALRVRLTKAGDG----AVSL 3774
Qy 3275 VVGDEAGVPVASVDRLELRPVDMGQLRA-VSVSAGRRGSLYAVQ----------WAE 3320
Db 3775 QAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGER-SLYRVEWQPRLLPAGPPRSWAV 3833
Qy 3321 VGP-------VPVCGQAWAWHEDVGESGGGPV---PGVVVLRCPDAGAGG---------- 3360

```
Db       3834 LGPDADRLAGTPGLGD---------QPDGGPTALYPEVRALR--KALAAGAPRPEAVVLP 3882

Qy       3361 --GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGAS 3418

Db       3883 VLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAVPGEEIGDLACAG 3942

Qy       3419 VWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGE 3478

Db       3943 VWGLVRSARSEHPGRFALVDTD---------------GHPDDRTAL-------PLALRAV 3980

Qy       3479 RLLAARLKRLESSGDVPAQRSGDTRA---RRSDVPAQRSGGVPARRSVDVSGREVLPWLS 3535

Db       3981 ---------LDGAGQL-SLRAGTARTPVLLRAGTPEEQRG--PA-------------FDP 4015

Qy       3536 GGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIV 3595

Db       4016 AGTVLVTGATGTLGRLLARHLAAEHGVRHLLLLSRGGRAAEGADELAAELAGLEAEPCFA 4075

Qy       3596 ACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 3655

Db       4076 ACDAADREALARVLAEVPADRPLTGVIHAAGVLDDGTLDALTPERIGTVMRPKADAALNL 4135

Qy       3656 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWE 3715

Db       4136 HELTRTSPLSVFAVFSGAAGILGRPGQANYAAANTFLDALAQHRRAHGLPAVSLAWGLWG 4195

Qy       3716 EASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALA--LDRPVLLPADLRPA----- 3768

Db       4196 GATGMTGHLSGTDLRRMRRSGIAPMTHDQGLALFDRALAASAEDPLLVPMRLDLAALVRE 4255

Qy       3769 ------PPLPPLLQDLLPATRR-RTTRTTTTGGAD---NGAQLHARLAGQTHEQQHTTLL 3818

Db       4256 RAEHGPDAVPGPLLGLLPARAAVRQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELV 4315

Qy       3819 ALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPN 3878

Db       4316 RLVRAEVSGVLGYSGPDAVEPGRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPT 4375

Qy       3879 PTTLTHHLHTQLQPQPD--NAVAPVLAELDKLESALSALDKTDSASE----RVTLRLKSL 3932

Db       4376 SQAVAEYLAAELAGPRDGGDTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAAL 4435

Qy       3933 MLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDL 3969

Db       4436 TPQGRNPSAPAPAPSDLDERLDSANDDDLFAFIEEQL 4472


RESULT 8
Q93HJ5
ID   Q93HJ5      PRELIMINARY;      PRT;  6146 AA.
AC   Q93HJ5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69192.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR002155; Thiolase.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF00108; thiolase; 1.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; UNKNOWN_1.
DR   PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   6146 AA;  637460 MW;  1D56F1108D458D00 CRC64;

  Query Match           36.1%;  Score 7516.5;  DB 2;  Length 6146;
  Best Local Similarity  37.8%;  Pred. No. 2.8e-314;
  Matches 1849;  Conservative  469;  Mismatches 1395;  Indels 1181;  Gaps   86;

Qy         4 MDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAA 63

Db      1504 LDALAEGRTPAGVVV------EGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFDEVCAA 1557

Qy        64 FAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGE 123

Db      1558 FDGLLPGALREVVF---GGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGHSIGE 1614

Qy       124 IAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARL 183

Db      1615 LAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDEL---------PDLPGGV 1666

Qy       184 AVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLL 243

Db      1667 SVAAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA 1726

Qy       244 PITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSP 303

Db      1727 AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLELGP 1786

Qy       304 HPVLT---------------------------------------------------------- 308

Db      1787 DGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFAPYGA 1846

Qy       309 --------------MGLQELAP------------------------------------- 316

Db      1847 RPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAVVVPASG 1906

Qy       317 ---------------------------------DLGDT---------------------- 321

Db      1907 SYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVSLLGTAGSL 1966

Qy       322 -TGTADTVI---------------------------------------MG---------- 331

Db      1967 SSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGMGRVAALEHPE 2026

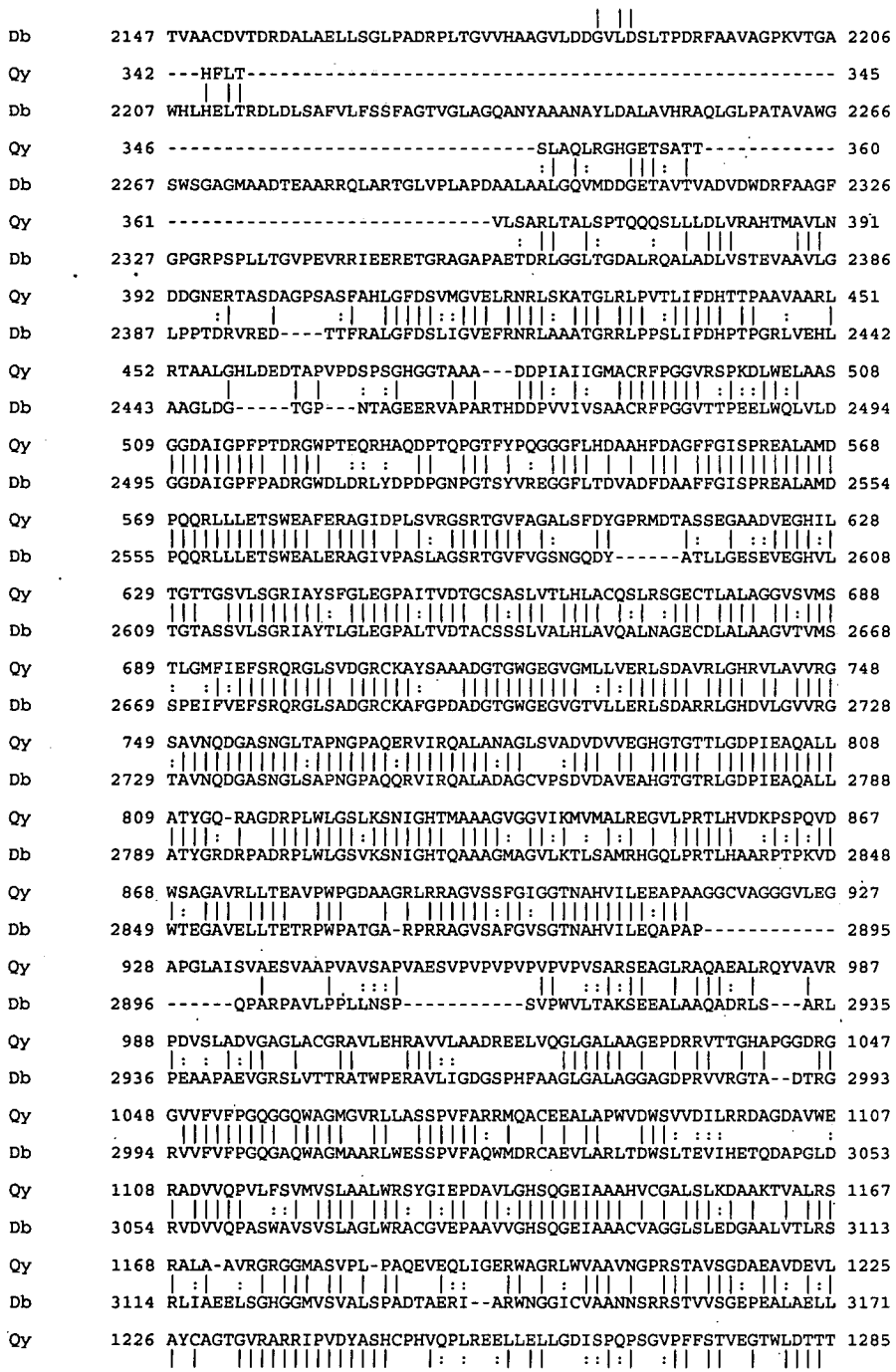
Qy       332 ---------------------------------------------------------- 331

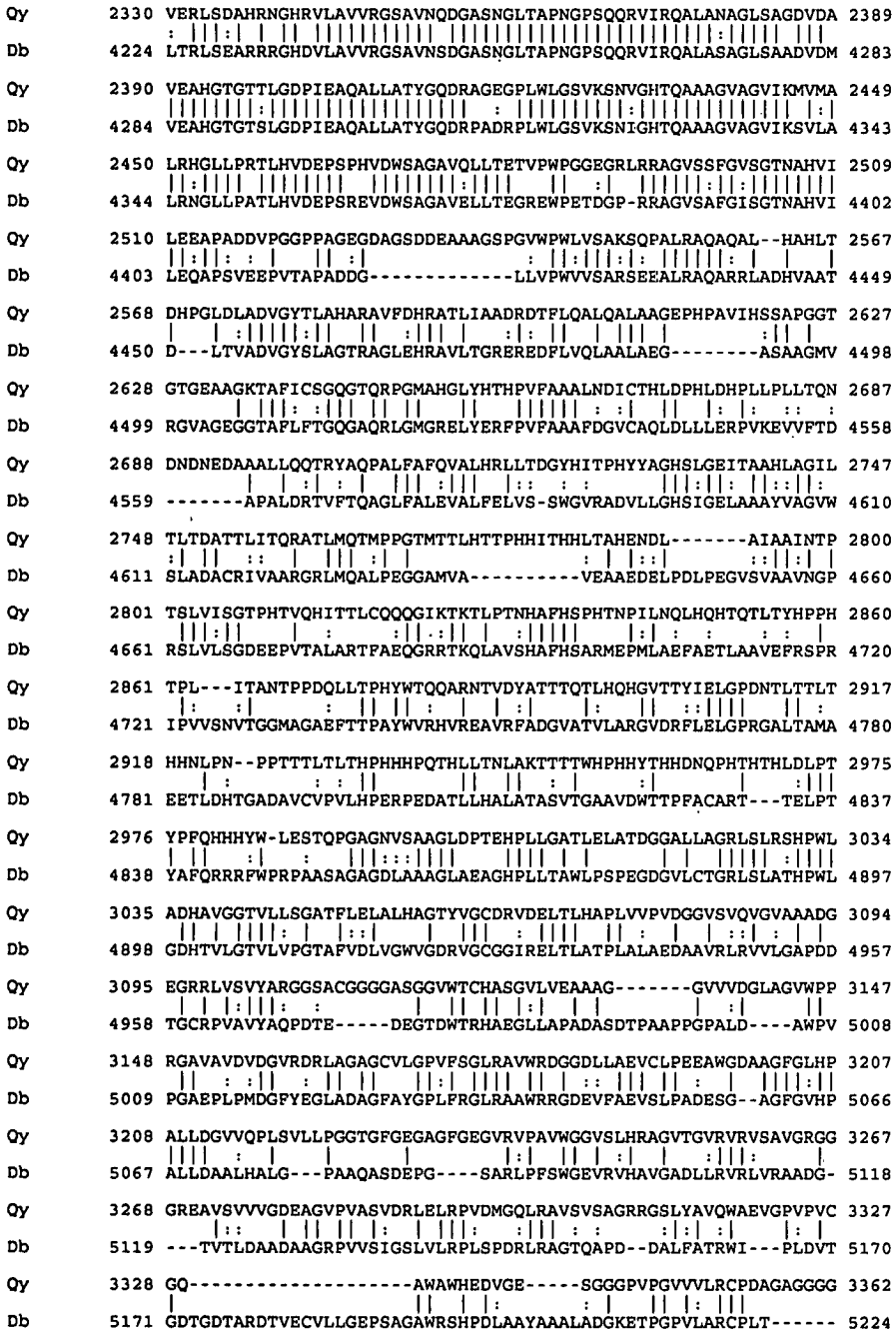
Db      2027 RWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRAGTRADAAPAGPS 2086

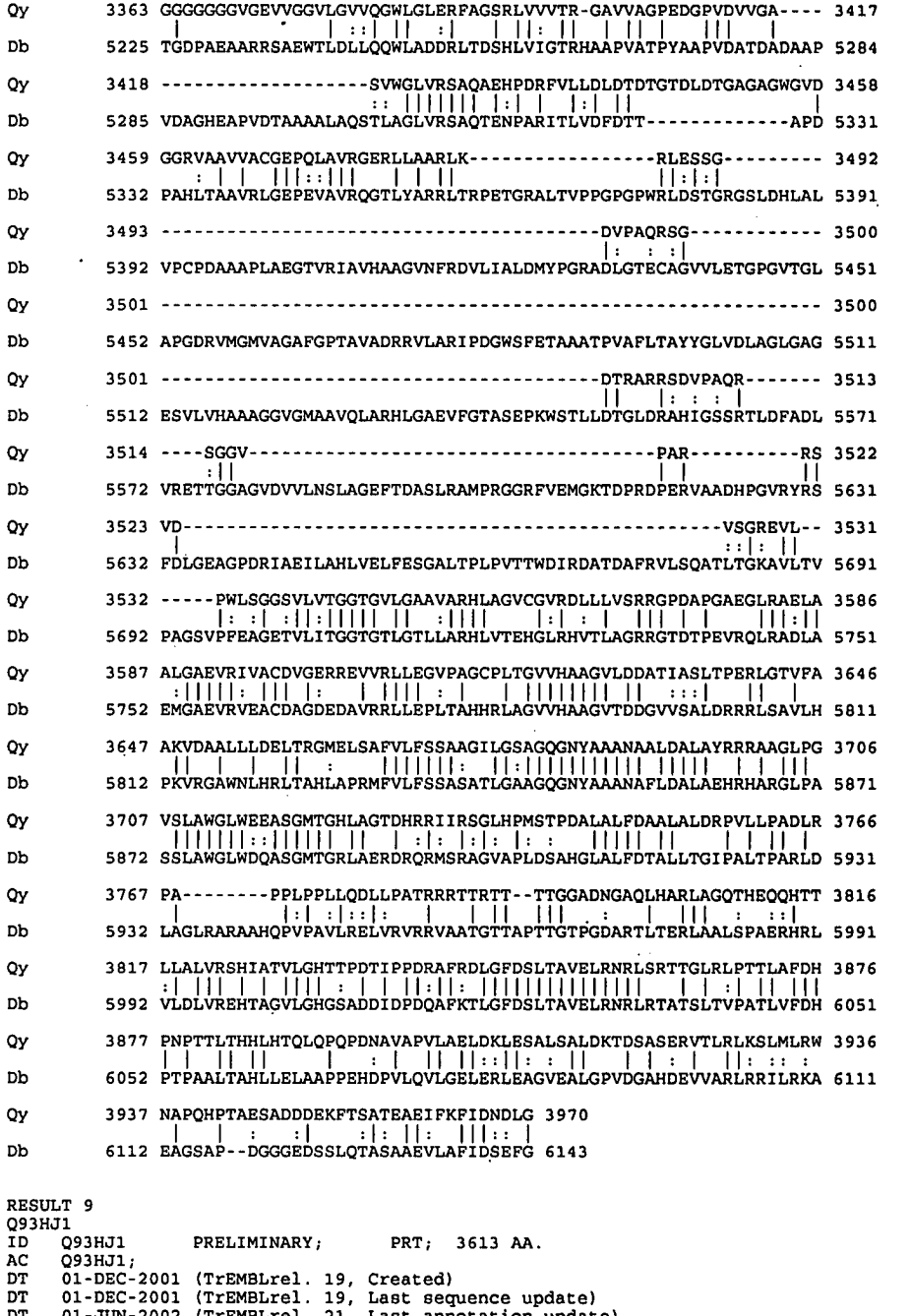
Qy       332 ---------------------------------TLRRG---------------------- 336

Db      2087 PWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGAAELTARLRETGTEV 2146

Qy       337 ---------------------------------QGTLD---------------------- 341
```


Db 2147 TVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSLTPDRFAAVAGPKVTGA 2206
Qy 342 ---HFLT---------- 345
Db 2207 WHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDALAVHRAQLGLPATAVAWG 2266
Qy 346 ----------SLAQLRGHGETSATT---------- 360
Db 2267 SWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMDDGETAVTVADVDWDRFAAGF 2326
Qy 361 ----------VLSARLTALSPTQQQSLLLDLVRAHTMAVLN 391
Db 2327 GPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGLTGDALRQALADLVSTEVAAVLG 2386
Qy 392 DDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARL 451
Db 2387 LPPTDRVRED----TTFRALGFDSLIGVEFRNRLAAATGRRLPPSLIFDHPTPGRLVEHL 2442
Qy 452 RTAALGHLDEDTAPVPDSPSGHGGTAAA---DDPIAIIGMACRFPGGVRSPKDLWELAAS 508
Db 2443 AAGLDG-----TGP---NTAGEERVAPARTHDDPVVIVSAACRFPGGVTTPEELWQLVLD 2494
Qy 509 GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMD 568
Db 2495 GGDAIGPFPADRGWDLDRLYDPDPGNPGTSYVREGGFLTDVADFDAAFFGISPREALAMD 2554
Qy 569 PQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHIL 628
Db 2555 PQQRLLLETSWEALERAGIVPASLAGSRTGVFVGSNGQDY------ATLLGESEVEGHVL 2608
Qy 629 TGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVMS 688
Db 2609 TGTASSVLSGRIAYTLGLEGPALTVDTACSSSLVALHLAVQALNAGECDLALAAGVTVMS 2668
Qy 689 TLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRG 748
Db 2669 SPEIFVEFSRQRGLSADGRCKAFGPDADGTGWGEGVGTVLLERLSDARRLGHDVLGVVRG 2728
Qy 749 SAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQALL 808
Db 2729 TAVNQDGASNGLSAPNGPAQQRVIRQALADAGCVPSDVDAVEAHGTGTRLGDPIEAQALL 2788
Qy 809 ATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVD 867
Db 2789 ATYGRDRPADRPLWLGSVKSNIGHTQAAAGMAGVLKTLSAMRHGQLPRTLHAARPTPKVD 2848
Qy 868 WSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEG 927
Db 2849 WTEGAVELLTETRPWPATGA-RPRRAGVSAFGVSGTNAHVILEQAPAP---------- 2895
Qy 928 APGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVR 987
Db 2896 ------QPARPAVLPPLLNSP----------SVPWVLTAKSEEALAAQADRLS---ARL 2935
Qy 988 PDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGHAPGGDRG 1047
Db 2936 PEAAPAEVGRSLVTTRATWPERAVLIGDGSPHFAAGLGALAGGAGDPRVVRGTA--DTRG 2993
Qy 1048 GVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWE 1107
Db 2994 RVVFVFPGQGAQWAGMAARLWESSPVFAQWMDRCAEVLARLTDWSLTEVIHETQDAPGLD 3053
Qy 1108 RADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRS 1167
Db 3054 RVDVVQPASWAVSVSLAGLWRACGVEPAAVVGHSQGEIAAACVAGGLSLEDGAALVTLRS 3113
Qy 1168 RALA-AVRGRGGMASVPL-PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVL 1225
Db 3114 RLIAEELSGHGGMVSVALSPADTAERI--ARWNGGICVAANNSRRSTVVSGEPEALAELL 3171
Qy 1226 AYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT 1285
Db 3172 AECEAEGVRARRIPVDYASHSAQVERIERQLTELAAGVAPRPGDIPFHSTTTGGPLDTTG 3231
Qy 1286 LDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGS 1345
Db 3232 LDADYWYRNLRRPVLFGPVTEELLAQGHDVFLEMSPHPVLVPAVQDAVDATGTTAAAVGS 3291
Qy 1346 LRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHT---------HLDLPTY 1396
Db 3292 LRRDDGGPERFTLSLAEAFVR--GAPVDW----------TAVLGTSGARRVELPTY 3335
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 3336 PFQRGRFWPEPRT----------AAATTGG-AEEDAPLWRAVDRGDLAAVAA 3376
Qy 1457 TLAVPPSAG---LDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLI- 1512
Db 3377 ELALPDGESLHDLHDLVPALSGWRRRRKDSALVDSWRYRVSWTPVTAPVLAALTGDWLLV 3436
Qy 1513 -AIPETQTHH-------PHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQA 1564
Db 3437 TADPDGEAARWARTALGASVTAQLTSPD----GPIP---------- 3468
Qy 1565 QNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATT 1624
Db 3469 ----AGPWTGVVSLLGLADRPHPDHPVVPVGVADTVALLSALREAGIDAPLWGLTRGAVG 3524
Qy 1625 THPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQ 1684
Db 3525 TGRSDQVRAPGQAQLWGLGRVAGLENPAAWGGLVDLPAELDDRAAGLLRAALAAGEDEQE 3584
Qy 1685 LAIRTTGTHTRRL--TPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQH 1742
Db 3585 YALRPAGLYVRRLVRAPLDGTAARRSWRPRPDGTVVVTGGTGALGARVARWL-AHAGAGH 3643
Qy 1743 LLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHT 1802
Db 3644 LLLTSRSGMAADGADDLVAELRALGARVTVAACDVADRAQLAAALDTVPDEFPVAAVVHT 3703
Qy 1803 AGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQAN 1862
Db 3704 AGVNRTATLDRTTLAELADVMSAKAAGARNLDELLAGHDLDAFVLFSSGAAIWGSAGQGA 3763
Qy 1863 YAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATA 1922
Db 3764 YAAANAYADALAADRRRRGLTATSVAWGSWAGGGMVDDDLARD-LARGGVRSMDPDRAIG 3822
Qy 1923 AVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARER-------EPAVQRPTPPA-- 1973
Db 3823 ALQQAL-DHDETTLTVTDMDWERFAGT------FTAARPRPLIDAIPEAARAATTDPAAG 3875
Qy 1974 --ELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFR 2031
Db 3876 ESELAARLAGSTDGERERELLDLVRTAAALALGHRGAEAITPSKPFKDLGFDSLTAVDLR 3935
Qy 2032 NLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAA--ESAPSAAAVTAEASVTEPIA 2089
Db 3936 NRLTAATGVRLPATLVFDHPTPLAAAVELRTRLFGGDALPTSAGSVAPTAAEAD--DPVV 3993
Qy 2090 IVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGG 2149
Db 3994 IVAMACRFPGEATSPEKLWDLIAAGKDGIGDFPADRGWSVAA---------DAAFSQTGG 4044
Qy 2150 FLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTN 2209
Db 4045 FLPDVAEFDAGFFGISPREALAMDPQQRLLLETSWEALERAGVDALKLRGSRTGVFVGAG 4104
Qy 2210 GQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLAC 2269
Db 4105 SHDYGTLLTSLEGGQD-YALTGAVGSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAA 4163
Qy 2270 QALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLL 2329
Db 4164 QALRGGECDVALAGGVAVMATPDAFDAFARQGGLAPDGRCKAFADGADGTGWGEGVGVLV 4223





RESULT 9
Q93HJ1
ID Q93HJ1 PRELIMINARY; PRT; 3613 AA.
AC Q93HJ1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Modular polyketide synthase.

```
GN   OLMA7.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69196.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3613 AA;  379236 MW;  6FAAE5CAB158E442 CRC64;

  Query Match             35.3%;  Score 7356;  DB 2;  Length 3613;
  Best Local Similarity   46.9%;  Pred. No. 1.1e-307;
  Matches 1646;  Conservative  413;  Mismatches 1141;  Indels  308;  Gaps   65;

Qy        479 ADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTF 538
              | :|:|::||||| |||| ||: ||::   | | :  || ||||          || :|||
Db         31 AREPLAVVGMACRLPGGVDSPEQLWQMVMDGADGVSGFPDDRGWDLSSLFDSDPDRPGTT 90

Qy        539 YPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTG 598
              | : | ||  |  |||||||||||||| |||||||||||||| ||||||| ::||||||
Db         91 YTREGAFLRGAGDFDAGFFGISPREALTMDPQQRLLLETSWEALERAGIDPHTLRGSRTG 150

Qy        599 VFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCS 658
              || |  : ::  ::  : ::     :|: :|| : ||:|||::|  ||||||:|||| ||
Db        151 VFVGGTAIEHTVKLMNSPTD-----QGYAITGGSASVMSGRVSYVLGLEGPALTVDTACS 205

Qy        659 ASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGT 718
              :||| || | |||| |:|:|||||||:||:|   |: |:|||||: ||||||:|  |||
Db        206 SSLVALHSAAQSLRQGDCSLALAGGVAVMATSSAFVTFARQRGLAADGRCKAFSDDADGI 265

Qy        719 GWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALAN 778
              |||||  ::|:|||:|| | || ||||:||||||||||||||:|||||:|:|||||:||
Db        266 GWGEGAAVVLLERLADARRNGHPVLAVIRGSAVNQDGASNGLSAPNGPSQQRVIRQAVAN 325

Qy        779 AGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAG 837
              |||:|||||:|| ||||||:|||||||||||||| |   |||||||||||| || | :|
Db        326 AGLALADVDMVEAHGTGTSLGDPIEAQALLATYGQDRHDGRPLWLGSLKSNIAHTQAVSG 385

Qy        838 VGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 897
              | |||| |:||| |||||:|||| |||| |||||||||| ||||    ||     | |||||||
Db        386 VAGVIKTVLALRNGVLPKTLHVGKPSTQVDWSAGAVELLTEGREWP--ETGGPRRAGVSS 443

Qy        898 FGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 957
              ||| |||||||||:||                      |  | |  | ||  |
Db        444 FGISGTNAHVILEQAP----------------------EEDSAEPQPVDAP--EPPLTTS 479

Qy        958 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLA--- 1014
              |  ||  :| ::|| |||||  |: ::   | :  |||    |   |: | ||||: |
Db        480 PATVPWLISGQTEAALRAQAGRLQAHLDAHPGLGTADVAHSLLASRSKLAHRAVLFAGQG 539

Qy       1015 ADREELVQGLGALAAGEPDRRVTTGHAPG-----GDRG-GVVFVFPGQGGQWAGMGVRLL 1068
              |   :   || ||| |           ||      || | |||||||||| || ||| |
Db        540 ARSGDRPTGLTALADGL--------DVPGLVRGTGDAGTGVVFVFPGQGSQWVGMGRDLW 591

Qy       1069 ASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWR 1128
              :|||| ||: | :||:|:|||||: ::: ||||| :| | |||||||::|||||||:||
Db        592 DASPVFRARMEECGQALSPYVDWSLRNVVFRDAGDPLWSRVDVVQPVLWAVMVSLAAVWR 651

Qy       1129 SYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR-ALAAVRGRGGMASVPLPAQ 1187
              |:|:|| ||:||||||:||| | | |||:| |: ||:|||  |  : |:||| || || :
Db        652 SFGVEPAAVVGHSQGEVAAACVAGGLSLEDGARVVAVRSRLVLEKLSGKGGMVSVALPVE 711

Qy       1188 EVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCP 1247
              ||: :  |: ||: |||||||  |  |||: :|:||:|| |  :|||||| ||||||
Db        712 GVEERLA-RFEGRIGVAAVNGPGSAVVSGEPDALDELLAECEESGVRARRIAVDYASHSA 770

Qy       1248 HVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQA 1307
              |  | ::||  | || |: | | |:||| |  |||: ||| ||  || : |||   |:
Db        771 QVDALNDDLLAELADIRPKSSAVAFYSTVTGERLDTSGLDARYWVTNLRERVRFEPVVRL 830

Qy       1308 LADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTG 1367
              ||:  ::|||| |||| |  ||::| |      ||:||||| :    ||||:||  :
Db        831 LAEQEYQVFVESSPHPVLTVAIQETQEGVFGTGTAVGSLRRDEGGADRFLTSLAEAYVA- 889

Qy       1368 IGTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLES-SQPGAGSGSGAGAGSGAGS 1426
              | |  |  :         |  :||||| ||| ||:|   |||
Db        890 -GAPVDWSVLFA-----GMPVRRVDLPTYAFQHERYWIEDVVAPGAD------------- 930

Qy       1427 GRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARI 1486
                    |   |:: || ||   ||: :|  :        : ||||||||  | : ::: :
Db        931 ------GVVDPVDAAFWGAVENADLQGMAALVDGAEPEVWEPVVPALSAWRRGRQERSVL 984

Qy       1487 NTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIP---LTL 1543
              ::| |:  |: :|:|:  :  |||:  |                  |  |:      |
Db        985 DSWRYRTVWRSVTVPSVGRLSGTWLVVTP-----------------GGAGPVEEVRRAL 1026

Qy       1544 NHTHTNPQHLHHTLHHTRQQAQNHTTGA--ITGLLSLLALDETPHPHHPHTPTGTLLNLT 1601
                        |       |:      |||  ::|::|||  ||   |       : :|
Db       1027 EAAGAEVVVLAWEAEEGREALAGRLTGAGEVSGVVSLLGWDEEAAP---------VATVT 1077

Qy       1602 LTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLP 1661
              | |         |||  | | :   | : || | | | | :   || |    |::|:|
Db       1078 LVQALADAGVEAPLWVLTQGAASVGAED-VVHPVQTQVWALGQVAGLEQPGSWGGLVDVP 1136

Qy       1662 TTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTP----TPHGTT 1717
                    || |   : |:|:|::| : |||      |   | |    ||||
Db       1137 GVWDERVASGLTAVLAAGEGEDQVAVRSSGAYARRL---VRAPLGANPAPVRDWNPHGTV 1193

Qy       1718 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 1777
              |||||| :  ||   |   :  : ||| ||:|     |  | |:|   :|   ||
Db       1194 LITGGTGGIGAHLARWL-AKEGAERLLLVSRSGEQAEGAAELATELGGLGAEVTFAACDV 1252

Qy       1778 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLT 1837
              |: | |  ::  || :|||| | ||||:   | |   ||      || ||   | |  ||
Db       1253 SDRDALGHVIAGIPAEHPLTAVFHTAGVSGYAELATATPEHYETVLSAKTRGARNLDALT 1312

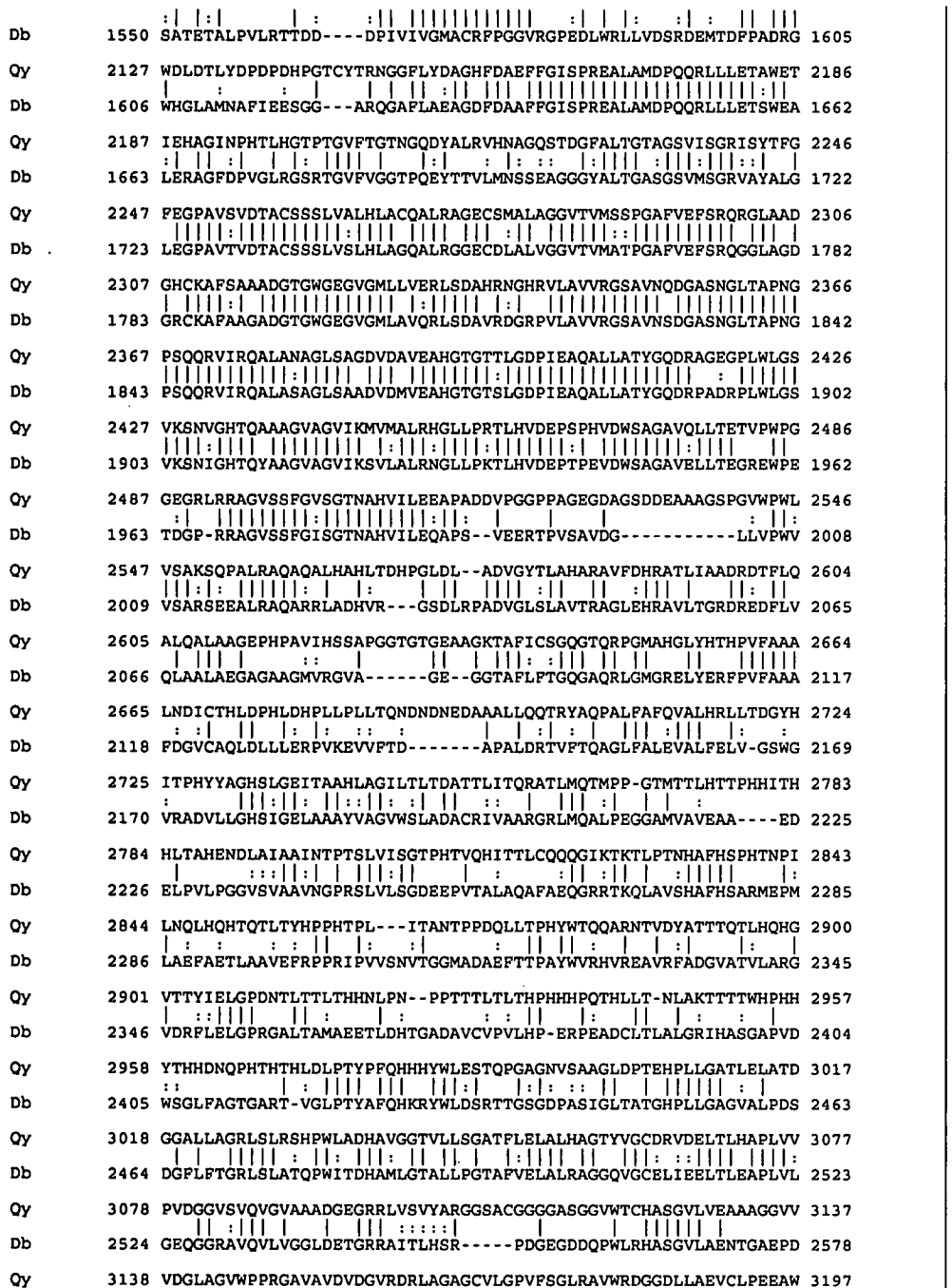
Qy       1838 QHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGL 1897
                | ||||:|| || :|: |   |||||||| || :|    | |||::|| |:  |:
Db       1313 AELTLEAFVLFSSGAAVWGSAGTGAYAAANAYLDGLAWNRRARGLVATSVSWGGWKATGM 1372

Qy       1898 ADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKI-------EHTS 1950
              | :|     | ||| | | | ||  |: ||:   |   : : |:||::
Db       1373 A-ADGTEEQLARRGVRAMEPALAIKALRQALEQDETA-LTVTDMDWARFTPGYTIARRRP 1430

Qy       1951 QTSDLVSAAR----EREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHA 2006
                |:   ||    : |||    |  | | :|||  |: :|   ||||||  | || |
Db       1431 LIEDIPEVARALSEDSEPAPDDGTDSA-LRQTLAGLTAPEQHDRLLELVRSEAATVLTHR 1489

Qy       2007 DPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG 2066
                |   : || ||||||||:| || |  |||||| :|||||||  :|| ||  :|
Db       1490 TTDDITAGRPFRDLGFDSLTAMELRNRLNTATGLRLPATLVFDHPTPQRLAGHLHEKLFD 1549

Qy       2067 TAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRG 2126
```



RESULT 10
Q9ALM2
ID Q9ALM2 PRELIMINARY; PRT; 5588 AA.
AC Q9ALM2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Polyketide synthase extender modules 8-10.
GN SPNE.
OS Saccharopolyspora spinosa.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=60894;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21257765; PubMed=11358695;
RA Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,

```
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23262.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5588 AA;  585629 MW;  C47F38F44BAC6A11 CRC64;

  Query Match             35.0%;  Score 7281.5;  DB 2;  Length 5588;
  Best Local Similarity   36.1%;  Pred. No. 3.1e-304;
  Matches 1821;  Conservative  490;  Mismatches 1349;  Indels 1389;  Gaps   86;

Qy         12 PAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWS 71
Db        546 PAANVVQGAAD-CKGKVAFVFPGQGSHWQGMARELSESSPVFRRKLAECAAATAPYVDWS 604

Qy         72 VEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSG 131
Db        605 LLGVLRGDPDAPALDRDDVIQLALFAMMVSLAELWRSCGVEPAAVVGHSQGEIAAAHVAG 664

Qy        132 GLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGP 191
Db        665 ALSLTDAVRIIAARCDAVSALTGKGGMLAIALPESAVVKRIAGLPE-----LTVAAVNGP 719

Qy        192 RSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSR 251
Db        720 GSTVVSGEPSALERLQTELTAENVQTRRVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAE 779

Qy        252 IPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGL 311
Db        780 IAFYSTVTGERTDTGRLDADYWYQNLRQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGI 839

Qy        312 QELAPDLGDTTGTADT--VIMGTLRRGQGTLDHFLTSLA--QLRG--------------- 352
Db        840 QE-------TLEAADAGGVVVGSLRRGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGAR 892

Qy        353 ------------------------------------------------------------ 352
Db        893 RVPLPTYPFQRQRYWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVK 952

Qy        353 ------------------------------------------------------------ 352
Db        953 TQPWLADHVVGGAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAV 1012

Qy        353 ------------------------------------------------------------ 352
Db       1013 GGPDEAGRRSVRVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPL 1072

Qy        353 ----------------------------------------HGE----------------- 355
Db       1073 DSFYARAAERGFDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQ 1132

Qy        356 ------------------------------------------------TSATTVLS--- 363
Db       1133 ALGAAEEDPDEGWLPFAWQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSIDS 1192

Qy        364 ------------------------------------------------------------ 363

Db       1193 LVLRQISDKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALGTEL 1252

Qy        364 ------------------------------------------------------------ 363

Db       1253 HPDLTGLADPPPDVVVVPCGASRQDLDVASEARAATQRMLDLIQDWLAAARFAGSRLVVV 1312

Qy        364 ------------------------------------------------------------ 363

Db       1313 TCGAASTGPAEGVSDLVHAASWGLLRSAQSENPDRFVLVDVDGTAESWRALAAAVRSGEP 1372

Qy        364 ------------------------------------------------------------ 363

Db       1373 QLALRAGEVRVPRLARCVAAEDSRIPVPGADGTVLISGGTGLLGGLVARHLVAERGVRRL 1432

Qy        364 ------------------------------------------------------------ 363

Db       1433 VLAGRRGWSAPGVTDLVDELVGLGAAVEVASCDVGDRAQLDRLLTTISAEFPLRGVVHAA 1492

Qy        364 ------------------------------------------------------------ 363

Db       1493 GALADGVVESLTPEHVAKVFGPKAAGAWHLHELTLDLDLSFFVLFSSFSGVAGAAGQGNY 1552

Qy        364 ------------------------------------------------------------ 363

Db       1553 AAANAFLDGLAQHRRTAGLPAVSLAWGLWEQPSGMTGALDAAGRSRIARTNPPMSAPDGL 1612

Qy        364 ------------------------------------------------------------ 363

Db       1613 RLFEMAFRVPGESLLVPVHVDLNALRADAADGGVPALLRDLVPAPVRRSAVNESADVNGL 1672

Qy        364 -ARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELR 422
Db       1673 VGRLRRLPDLDQETQLLGLVREHVSAVLGHSG----AVEVGADRAFRDLGFDSLSGVEFR 1728

Qy        423 NRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDP 482
Db       1729 NRLGGVLGVRLPATAVFDYPTPRALVRFLLDKLIGGVEAPT-PAPAAVA----AVTADDP 1783

Qy        483 IAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQG 542
Db       1784 VVIVGMGCRYPGGVSSPEELWRLVAGGLDAVAEFPDDRGWDQAGLFDPDPDRLGTSYVCE 1843

Qy        543 GGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAG 602
Db       1844 GGFLRDAAEFDAGFFGISPREALAMDPQQRLLLEVAWETVERAGIDPLSLRGSRTGVFAG 1903

Qy        603 ALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLV 662
Db       1904 LMHHDYGARFITRAPEG---FEGYLGNGSAGGVFSGRVAYSFGFEGPAVTVDTACSSSLV 1960

Qy        663 TLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGE 722
Db       1961 ALHLAGQALRSGECDLALAGGVTVMATPGMFVEFSRQRGLAADGRCKSFAAAADGTGWGE 2020

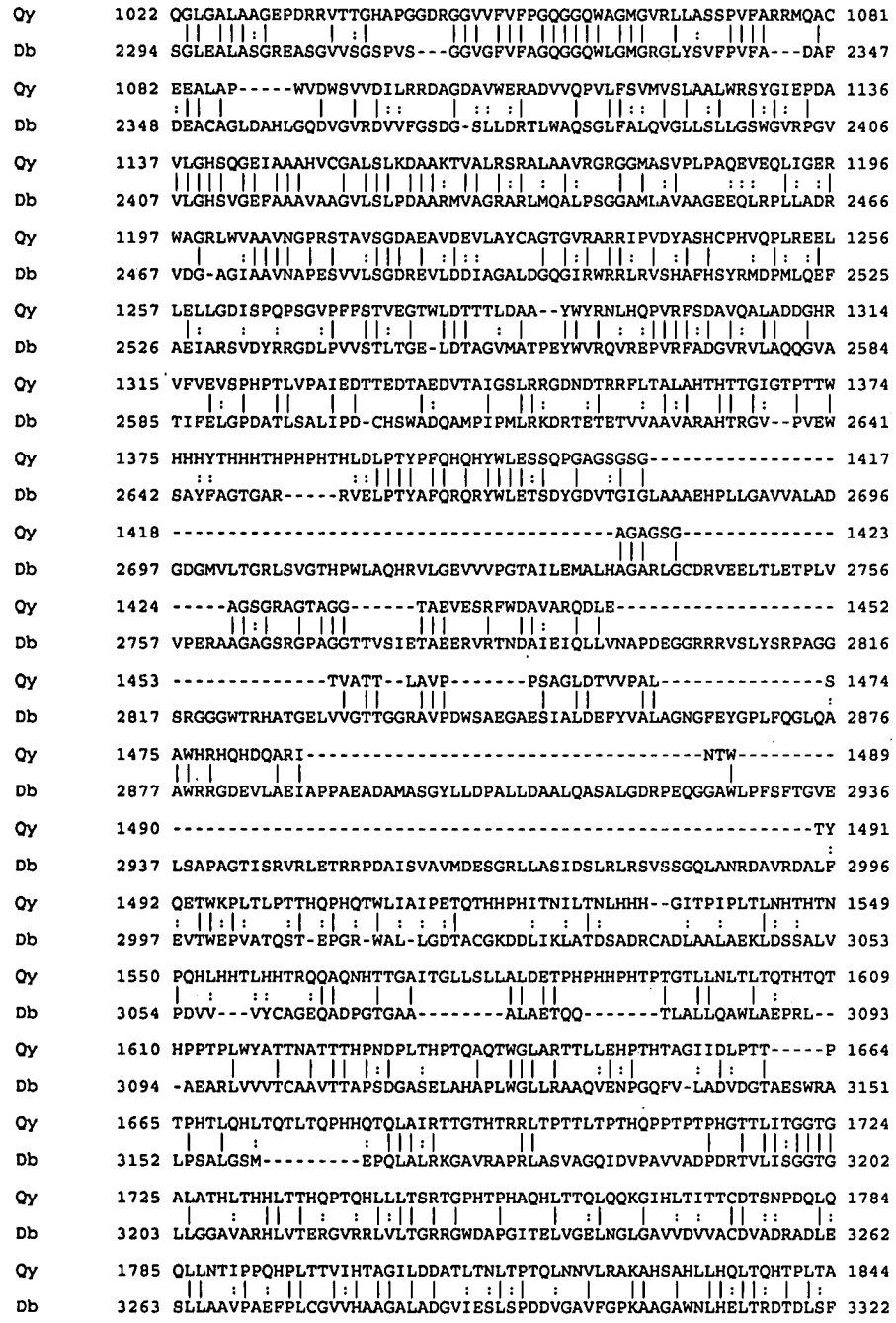
Qy        723 GVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLS 782
Db       2021 GAGLVLLERLSDARRNGHAVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALASAGLS 2080

Qy        783 VADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGV 841
Db       2081 VSDVDAVEAHGTGTRLGDPIEAQALIATYGQGRDSDRPLWLGSVKSNIGHTQAAAGVAGV 2140

Qy        842 IKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIG 901
Db       2141 IKMVMAMRHGQLPATLHVDEPTSEVDWSAGDVQLLTENTPWPGNS--HPRRVGVSSFGIS 2198

Qy        902 GTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPV 961
Db       2199 GTNAHVILEQA----------------------SKTPDETADKSGPDSEST-VDLPA-V 2233

Qy        962 PVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELV 1021
Db       2234 PLIVSGRTPAALSAQASALLSYLGERGDISTLDAAFSLASSRAALEERAVVLGADRETLL 2293
```


Qy 1022 QGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQAC 1081
Db 2294 SGLEALASGREASGVVSGSPVS---GGVGFVFAGQGGQWLGMGRGLYSVFPVFA---DAF 2347
Qy 1082 EEALAP-----WVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDA 1136
Db 2348 DEACAGLDAHLGQDVGVRDVVFGSDG-SLLDRTLWAQSGLFALQVGLLSLLGSWGVRPGV 2406
Qy 1137 VLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGER 1196
Db 2407 VLGHSVGEFAAAVAAGVLSLPDAARMVAGRARLMQALPSGGAMLAVAAGEEQLRPLLADR 2466
Qy 1197 WAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREEL 1256
Db 2467 VDG-AGIAAVNAPESVVLSGDREVLDDIAGALDGQGIRWRRLRVSHAFHSYRMDPMLQEF 2525
Qy 1257 LELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA--YWYRNLHQPVRFSDAVQALADDGHR 1314
Db 2526 AEIARSVDYRRGDLPVVSTLTGE-LDTAGVMATPEYWVRQVREPVRFADGVRVLAQQGVA 2584
Qy 1315 VFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTW 1374
Db 2585 TIFELGPDATLSALIPD-CHSWADQAMPIPMLRKDRTETETVVAAVARAHTRGV--PVEW 2641
Qy 1375 HHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSG---------- 1417
Db 2642 SAYFAGTGAR-----RVELPTYAFQRQRYWLETSDYGDVTGIGLAAAEHPLLGAVVALAD 2696
Qy 1418 ----------AGAGSG---------- 1423
Db 2697 GDGMVLTGRLSVGTHPWLAQHRVLGEVVVPGTAILEMALHAGARLGCDRVEELTLETPLV 2756
Qy 1424 -----AGSGRAGTAGG------TAEVESRFWDAVARQDLE---------- 1452
Db 2757 VPERAAGAGSRGPAGGTTVSIETAEERVRTNDAIEIQLLVNAPDEGGRRRVSLYSRPAGG 2816
Qy 1453 --------------TVATT--LAVP-------PSAGLDTVVPAL----------S 1474
Db 2817 SRGGGWTRHATGELVVGTTGGRAVPDWSAEGAESIALDEFYVALAGNGFEYGPLFQGLQA 2876
Qy 1475 AWHRHQHDQARI----------NTW--------- 1489
Db 2877 AWRRGDEVLAEIAPPAEADAMASGYLLDPALLDAALQASALGDRPEQGGAWLPFSFTGVE 2936
Qy 1490 ----------TY 1491
Db 2937 LSAPAGTISRVRLETRRPDAISVAVMDESGRLLASIDSLRLRSVSSGQLANRDAVRDALF 2996
Qy 1492 QETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHH--GITPIPLTLNHTHTN 1549
Db 2997 EVTWEPVATQST-EPGR-WAL-LGDTACGKDDLIKLATDSADRCADLAALAEKLDSSALV 3053
Qy 1550 PQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQT 1609
Db 3054 PDVV---VYCAGEQADPGTGAA--------ALAETQQ-------TLALLQAWLAEPRL-- 3093
Qy 1610 HPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTT-----P 1664
Db 3094 -AEARLVVVTCAAVTTAPSDGASELAHAPLWGLLRAAQVENPGQFV-LADVDGTAESWRA 3151
Qy 1665 TPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTG 1724
Db 3152 LPSALGSM---------EPQLALRKGAVRAPRLASVAGQIDVPAVVADPDRTVLISGGTG 3202
Qy 1725 ALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ 1784
Db 3203 LLGGAVARHLVTERGVRRLVLTGRRGWDAPGITELVGELNGLGAVVDVVACDVADRADLE 3262
Qy 1785 QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTA 1844
Db 3263 SLLAAVPAEFPLCGVVHAAGALADGVIESLSPDDVGAVFGPKAAGAWNLHELTRDTDLSF 3322
Qy 1845 FVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNGLADSDKA 1903
Db 3323 FALFSSLSGVAGAPGQGNYAAANAFLDALAHYRRSQGLPAVSLAWGLWEQPSGMTET--- 3379
Qy 1904 RAYLDRRGFRPMSPELAT---AAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAAR 1960
Db 3380 LSEVDRSRIARANPPLSTKEGLRLFDAGLALDRAAVVPAKLDRTFLAEQARSGSL----- 3434
Qy 1961 EREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRAL 2020
Db 3435 ---PALLTALVPPIRRNRRASGTELADEGTLLGVVREHAAAVLGYSSAADVGVERAFRDL 3491
Qy 2021 GFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAV-T 2079
Db 3492 GFDSLSGVELRNRLAGVLGVRLPATAVFDYPTPRALARFLHQEL---ADEIATTPAPVTT 3548
Qy 2080 AEASVTEP--IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 3549 TRAPVAEDDLVAIVGMGCRFPGQVSSPEELWRLVAGGVDAVADFPADRGWDLAGLFDPDP 3608
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 3609 ERAGKTYVREGAFLTDADRFDAGFFGISPREALAMDPQQRLLLELSWEAIERAGIDPGSL 3668
Qy 2198 HGTPTGVFTGTNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDT 2256
Db 3669 RGSRTGVFAGLMYHDYGARFASRAPEGFEGYLGNGSAGSVASGRIAYSFGFEGPAVTVDT 3728
Qy 2257 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2316
Db 3729 ACSSSLVALHLAGQSLRSGECDLALAGGVTVMSTPGTFVEFSRQRGLAPDGRCKSFAESA 3788
Qy 2317 DGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2376
Db 3789 DGTGWGEGAGLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIQQA 3848
Qy 2377 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2436
Db 3849 LASAGLSVSDVDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQA 3908
Qy 2437 AAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2496
Db 3909 AAGVAGVIKMVMAMRHGQLPRTLHVDAPSSQVDWSAGRVQLLTENTPWP-DSGRPCRVGV 3967
Qy 2497 SSFGVSGTNAHVILEEAPADDVPGGPPAGEGD-AGSDDEAAAGSPGVWPWLVSAKSQPAL 2555
Db 3968 SSFGISGTNAHVILEQS----------TGQMDQAAEPDSSPVLDVPVVPWVVSGKTPEAL 4017
Qy 2556 RAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPH 2615
Db 4018 SAQAATLATYLDQNVDVSPLDVGISLAVTRSALDERAVVLGSDRDTLLSGLNALAAG--- 4074
Qy 2616 PAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPH 2675
Db 4075 ----HEAAGVVTGPVGIGGRTGFVFAGQGGQWLGMGRRLYSEFPAFAGAFDEACAELDAN 4130
Qy 2676 LDHP--LLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGH 2733
Db 4131 LGREVGVRDVVFGSDE------SLLDRTLWAQSGLFALQVGLWELL-GTWGVRPSVVLGH 4183
Qy 2734 SLGEITAAHLAGILTLTDATTLITQRATLMQTMPP-GTMTTLHTTPHHITHHLTAHENDL 2792
Db 4184 SVGELAAAFAAGVLSMAEAARLVAGRARLMQALPSGGAMLAVSATEARVGPLLDGVRDRV 4243
Qy 2793 AIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQ 2852
Db 4244 GVAAVNAPGSVVLSGDRDVLDGIAGRLDGQGIRSRWLRVSHAFHSHRMDPMLAEFAELAR 4303
Qy 2853 TLTYHPPHTPLITANT----PPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELG 2908
Db 4304 SVDYRSPRLPIVSTLTGNLDDVGVMATPEYWVRQVREPVRFADGVQALVDQGVDTIVELG 4363
Qy 2909 PDNTLTTLTHHNLPNPPTTT-LTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHT 2967

```
Db       4364 PDGALSSLVQECVAESGRATGIPLVRRDRDEVRTVLDALAQT----------------HT 4407

Qy       2968 H----------------THLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATL 3012

Db       4408 RGGAVDWGSFFAGTRATQVDLPTYAFQRQRYWLEPSD--SGDVTGVGLTGAEHPLLGAVV 4465

Qy       3013 ELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLH 3072

Db       4466 PVAGGDEVLLTGRLSVGTHPWLAEHRVLGEVVVPGTALLEMAWRAGSQVGCERVEELTLE 4525

Qy       3073 APLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAA 3132

Db       4526 APLVLPERGAAAVQLAVGAPDEAGRRSLQLYSRGAD-------EDGDWRRIASGLLAQAN 4578

Qy       3133 AGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCL 3192

Db       4579 A---VPPADSTAWPPDGAGQVDLAEFYERLAERGLTYGPVFQGLRAAWRHGDDIFAELA- 4634

Qy       3193 PEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAG 3252

Db       4635 ---GSPDASGFGIHPALLDAALHAMAL----------GASPDSEARLPFSWRGAQLYRAE 4681

Qy       3253 VTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGS 3312

Db       4682 GAALRVRLSPLGSG--+-AVSLTLVDATGRRVAAVESLSTRPVSTDQIGA---GRGDQER 4734

Qy       3313 LYAVQW---AEVGPVPVCGQAWA------WHEDVGESG-----------------GGPV 3345

Db       4735 LLHVEWVRSAESAGMSLTSCAVVGLGEPEWHAALKTTGVQVESHADLASLATEVAKRGSA 4794

Qy       3346 PGVVVLRCPDAGAGGGGGGGGGGVGEV-------VGGVLGVVQGWLGLERFAGSRLVVV 3398

Db       4795 PGAVIVPCPRPRA----------MQELPTAARRATQQAMAMLQQWLADDRFVSTRLILL 4843

Qy       3399 TRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVD 3458

Db       4844 THRAVSAVAGEDVLDLVHAPLWGLVRSAQAEHPDRFALIDMDDERASQT---------- 4892

Qy       3459 GGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVP 3518

Db       4893 --ALAEALTAGEAQLAVRSGVVLAPRLGQ---------------------------- 4919

Qy       3519 ARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA 3578

Db       4920 ----VKVSGGEAFRW-DEGTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGA 4974

Qy       3579 EGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTP 3638

Db       4975 DELVAELEQAGADVAVVACDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSP 5034

Qy       3639 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 3698

Db       5035 ERLDAVLRPKVDAAWYLHELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHC 5094

Qy       3699 RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRP 3758

Db       5095 RQAQGLPALSLASGLWASIDGMAGDLAAADVERLSRAGIGPLSAPGGLALFDAAVGSDEP 5154

Qy       3759 VLLPADL----------RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQ 3808

Db       5155 LLAPVRLDVEALRVQARSVQTRIPEMLHGMAMGPSRRTPFTSRV------EPLHERLAGL 5208

Qy       3809 THEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRL 3868

Db       5209 SEGERRQQVLQRVRADIAVVLGHGRSSDVDIEKPLAELGFDSLTAIELRNRLATATGLRL 5268

Qy       3869 PTTLAFDHPNPTTLTHHLHTQL----QPQP-----DNAVAPVLAELDKLESALSALDKTD 3919

Db       5269 PATLAFDHGTAAALAQHVCAQLGTATAPAPRRTDDNDATEPVRSLFQQAYAAGRILDGMD 5328

Qy       3920 SASERVTLR 3928

Db       5329 LVKVAAQLR 5337
```

RESULT 11
O30764
ID O30764 PRELIMINARY; PRT; 4340 AA.
AC O30764;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Polyketide synthase modules 1 and 2.
GN NIDA1.
OS Streptomyces caelestis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=36816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL-2821;
RX MEDLINE=98053867; PubMed=9393718;
RA Kakavas S.J., Katz L., Stassi D.;
RT "Identification and characterization of the niddamycin polyketide
RT synthase genes from Streptomyces caelestis.";
RL J. Bacteriol. 179:7515-7522(1997).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AF016585; AAC46024.1; -.
DR InterPro; IPR002106; AAtRNA_ligaseII.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00106; adh_short; 2.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR TIGRFAMs; TIGR00128; fabD; 3.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Oxidoreductase; Phosphopantetheine; Transferase.
SQ SEQUENCE 4340 AA; 457589 MW; D59A734CB5FB795D CRC64;

Query Match 34.8%; Score 7244; DB 2; Length 4340;
Best Local Similarity 42.6%; Pred. No. 9.1e-303;
Matches 1764; Conservative 421; Mismatches 1438; Indels 520; Gaps 99;

```
Qy          7 GEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLD-------ASDVFRESVRA 59

Db        518 GQTPR---GVRIGSTD-ADGRLALLFTGQGAQHPGMGQELYTTDPHFAAALDEVCEELQR 573

Qy         60 C------EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEP 113

Db        574 CGTQNLREVMFTP----------DQPDL--LDRTEYTQPALFALQTALYRTLTARGTQA 620

Qy        114 CAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIA 173

Db        621 HLVLGHSVGEITAAHIAGVLDLPDAARLITARAHVMGQLPHGGAMLSVQAAEHDL----- 675

Qy        174 PWTEDNPAR---LAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPL 230

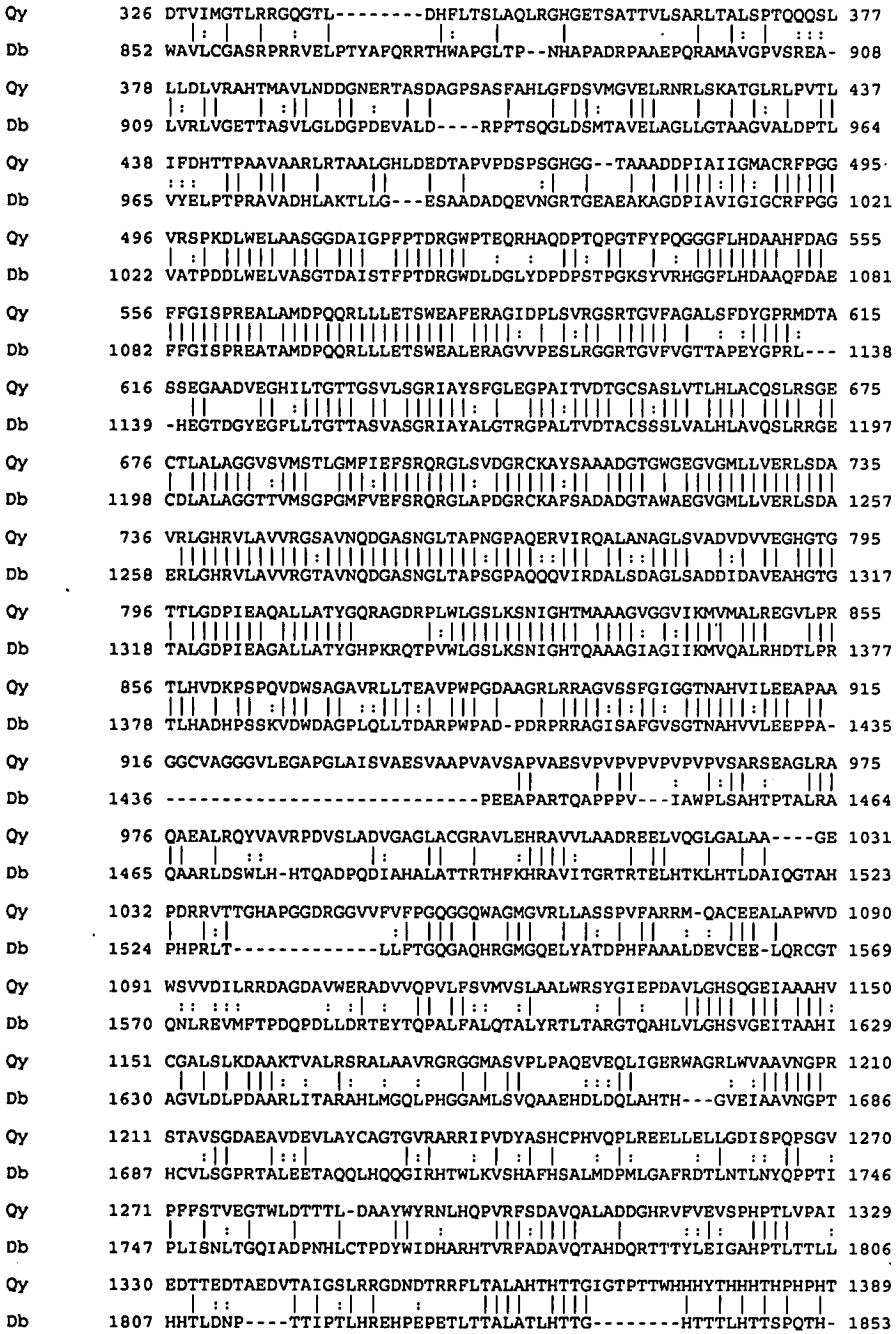
Db        676 ----DQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKVSHAFHSAL 731

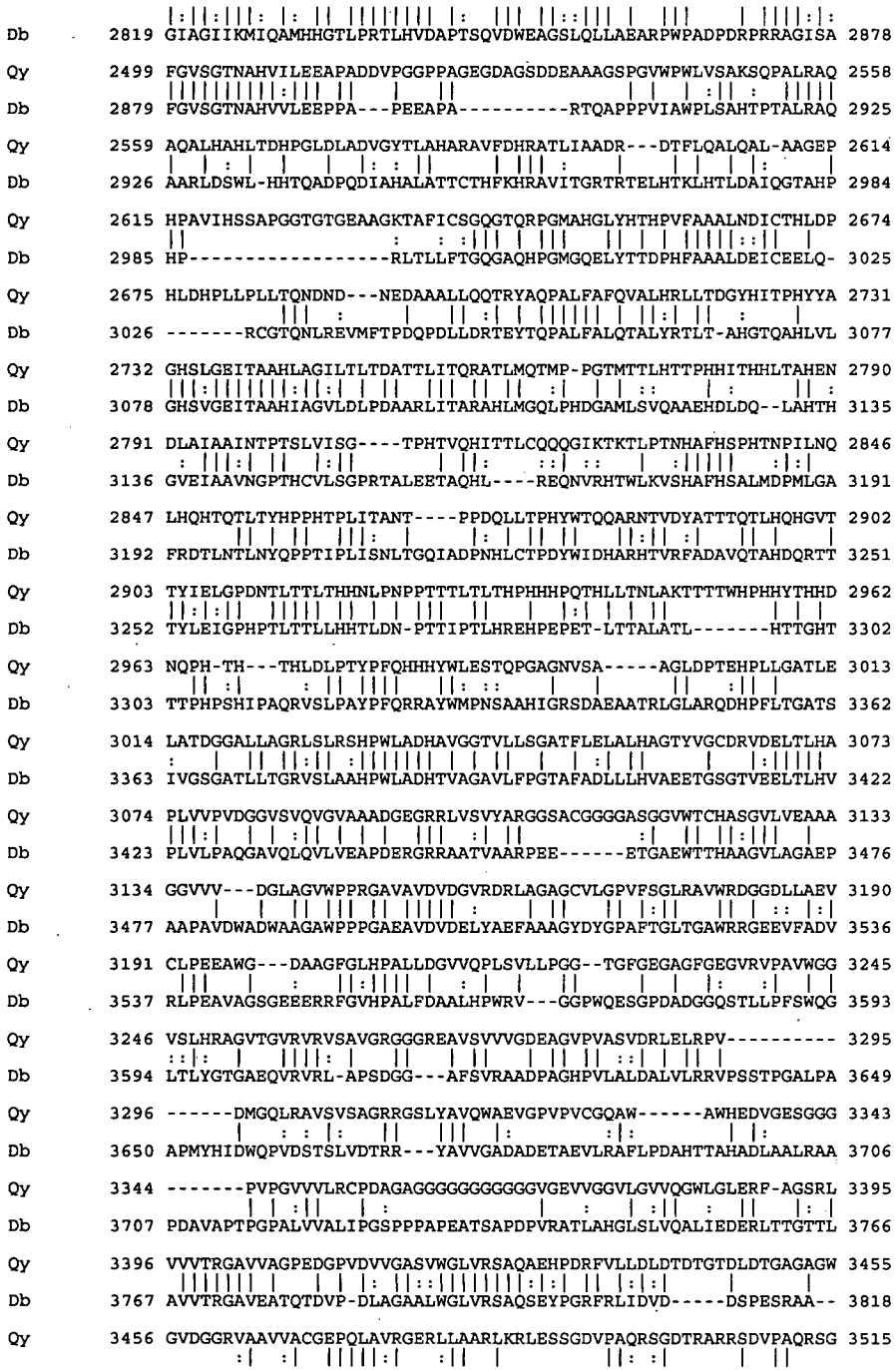
Qy        231 MYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL-DAAYWYRNMSSTVRFEPAARL 289

Db        732 MDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQT 791

Qy        290 LLQQGPKTFVEMSPHPVLTMGL----------------------QELAPDLGDTTGTA 325

Db        792 AHHQGTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHRERPEPETLTQAIAAVGVRTDGID 851
```


Qy 326 DTVIMGTLRRGQGTL--------DHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSL 377
Db 852 WAVLCGASRPRRVELPTYAFQRRTHWAPGLTP--NHAPADRPAAEPQRAMAVGPVSREA- 908
Qy 378 LLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTL 437
Db 909 LVRLVGETTASVLGLDGPDEVALD----RPFTSQGLDSMTAVELAGLLGTAAGVALDPTL 964
Qy 438 IFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGG--TAAADDPIAIIGMACRFPGG 495
Db 965 VYELPTPRAVADHLAKTLLG---ESAADADQEVNGRTGEAEAKAGDPIAVIGIGCRFPGG 1021
Qy 496 VRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAG 555
Db 1022 VATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQFDAE 1081
Qy 556 FFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTA 615
Db 1082 FFGISPREATAMDPQQRLLLETSWEALERAGVVPESLRGGRTGVFVGTTAPEYGPRL--- 1138
Qy 616 SSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGE 675
Db 1139 -HEGTDGYEGFLLTGTTASVASGRIAYALGTRGPALTVDTACSSSLVALHLAVQSLRRGE 1197
Qy 676 CTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDA 735
Db 1198 CDLALAGGTTVMSGPGMFVEFSRQRGLAPDGRCKAFSADADGTAWAEGVGMLLVERLSDA 1257
Qy 736 VRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTG 795
Db 1258 ERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDAVEAHGTG 1317
Qy 796 TTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPR 855
Db 1318 TALGDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGIIKMVQALRHDTLPR 1377
Qy 856 TLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAA 915
Db 1378 TLHADHPSSKVDWDAGPLQLLTDARPWPAD-PDRPRRAGISAFGVSGTNAHVVLEEPPA- 1435
Qy 916 GGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRA 975
Db 1436 ----------PEEAPARTQAPPPV---IAWPLSAHTPTALRA 1464
Qy 976 QAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAA----GE 1031
Db 1465 QAARLDSWLH-HTQADPQDIAHALATTRTHFKHRAVITGRTRTELHTKLHTLDAIQGTAH 1523
Qy 1032 PDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRM-QACEEALAPWVD 1090
Db 1524 PHPRLT----------LLFTGQGAQHRGMGQELYATDPHFAAALDEVCEE-LQRCGT 1569
Qy 1091 WSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHV 1150
Db 1570 QNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARGTQAHLVLGHSVGEITAAHI 1629
Qy 1151 CGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPR 1210
Db 1630 AGVLDLPDAARLITARAHLMGQLPHGGAMLSVQAAEHDLDQLAHTH---GVEIAAVNGPT 1686
Qy 1211 STAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGV 1270
Db 1687 HCVLSGPRTALEETAQQLHQQGIRHTWLKVSHAFHSALMDPMLGAFRDTLNTLNYQPPTI 1746
Qy 1271 PPFSTVEGTWLDTTTL-DAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAI 1329
Db 1747 PLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQTAHDQRTTTYLEIGAHPTLTTLL 1806
Qy 1330 EDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHT 1389
Db 1807 HHTLDNP----TTIPTLHREHPEPETLTTALATLHTTG--------HTTTLHTTSPQTH- 1853
Qy 1390 HLDLPTYPFQHQHYWLE-------SSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRF 1442
Db 1854 HLDLPTYPFQRDRYWMEPVRVAQVSGQPGAD----------RLRY 1888
Qy 1443 ---WDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLT 1499
Db 1889 RVVWEA----------APEDHTVPSAERW----------LL 1909
Qy 1500 LPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHH 1559
Db 1910 LGGSDPADAGILVGVEE----------QLVGHGAA---------- 1934
Qy 1560 TRQQAQNHTTGAITGLLSLLALDETP--HPHHPHTPTGTLLN-------LTLTQTHTQTH 1610
Db 1935 ----VQRLDIPAPAGRRLLPSCSELPVRGPSHTGEFCGSRLRSPRVAEAVALLQALGDAG 1990
Qy 1611 PPTPLWYATTNATTTHPNDPLTHPT--QAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT 1668
Db 1991 ADAPLWIATRGAVAAQAGEA---PSVGGAQLWGLGQVAGLELADRWGGLVDLPADPAPAA 2047
Qy 1669 LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALAT 1728
Db 2048 LRGLARVLVSNAPDNQVAIRASGVFVRRVVPAPGRPVRNDWAPS--GTVLITGGTGALGS 2105
Qy 1729 HLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLN 1788
Db 2106 QVARRLAL-AGAPHLLLAGRRG--NSGAAELVDELTALGAEVTVAACDAADRDALVSLLA 2162
Qy 1789 TIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLY 1848
Db 2163 TIPEHRPLTAVLHAAGVLDDGVLDGLTPERIDAVLRAKATAARHLDELTADLDLDAFVLF 2222
Qy 1849 SSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLD 1908
Db 2223 SSIVGVWGNGGQATYAAANAALDALAHGRRARGQRATSIAWGPWAGSGMAAGDGAKS-LA 2281
Qy 1909 RRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQR 1968
Db 2282 RDGIDVLEPEEALDLLEQAVRSDETT-LLVADVDWETFLSSSASRRFQLPLFGRIPAARA 2340
Qy 1969 PT------PPAELHKTLAHQTSADQR-AALLELVRDHVAAVLRHADPKAIAPDQSFRALG 2021
Db 2341 VAVSGADGSPDDVPVWLSPQLSAAERQLRLLDLVRTEAASVLRHGSATAVDAETAFRAGG 2400
Qy 2022 FDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGT---AAESAPSAAAV 2078
Db 2401 FDSLTLVELRNRLASATGLRLSNTLLFDHPTPAAVAEYLGEQLLGTEPSAAASHGDAGAR 2460
Qy 2079 TAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPD 2138
Db 2461 TVVAS-DEPVAVVGMACRYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPS 2519
Qy 2139 HPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLH 2198
Db 2520 TPGKSYVRHGGFLHDAAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLR 2579
Qy 2199 GTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTAC 2258
Db 2580 TTRTGVFTGASQQEYGTQSREAADKYGGHLLTGTLASVMSGRVAYTLGLQGPALTVDTAC 2639
Qy 2259 SSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADG 2318
Db 2640 SSSLVALHLAVQSLRRGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKPFADTADG 2699
Qy 2319 TGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALA 2378
Db 2700 TAWAEGVGVLLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALS 2759
Qy 2379 NAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAA 2438
Db 2760 DAGLSADDIDAVEAHGTGTPLGDPIEAEALLATYGHPER-QTPVWLGSLKSNIGHSQAAA 2818
Qy 2439 GVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSS 2498



RESULT 12
Q9X993
ID Q9X993 PRELIMINARY; PRT; 6797 AA.
AC Q9X993;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Polyketide synthase (PIMS1 protein).
GN PIMS1.
OS Streptomyces natalensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=68242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214571; PubMed=10187796;
RA Aparicio J.F., Colina A.J., Ceballos E., Martin J.F.;
RT "The biosynthetic gene cluster for the 26-membered ring polyene
RT macrolide pimaricin. A new polyketide synthase organization encoded by
RT two subclusters separated by functionalization genes.";
RL J. Biol. Chem. 274:10133-10139(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Aparicio J.F., Fouces R., Mendes M.V., Olivera N.;
RT "A complex multienzyme system encoded by five polyketide synthase
RT genes is involved in the biosynthesis of the 26-membered polyene
RT macrolide pimaricin in Streptomyces natalensis.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ132222; CAB41041.1; -.
DR EMBL; AJ278573; CAC20931.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR001899; Gram_pos_anchor.

```
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00626; RCC1_2; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6797 AA;  710149 MW;  1D2DF85074D050DA CRC64;

  Query Match             34.7%;  Score 7221.5;  DB 2;  Length 6797;
  Best Local Similarity   41.8%;  Pred. No. 1.5e-301;
  Matches 1784;  Conservative  453;  Mismatches 1373;  Indels  653;  Gaps  107;

Qy         52 VFRESVRACEAAFAPYVDWS-VEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQG 110
Db       2846 VTRGAVSGADLAGAAV--WGLVRSALSEHPGRFGL--VDLDDDAELALVPRVLA-----S 2896

Qy        111 VEP-CAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTL--AGTGALVSVAATPDE 167
Db       2897 DEPQLLVRG---GEVLAAR----LARAQSSHAVT-WDPSGTVLVTGGTGGLGRVMAR--H 2946

Qy        168 LLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADL------------------VA 208
Db       2947 LV------VEHGVRNLLLVSRRGP---AAEGAEELVTELRHSGAEVAVEACDVTDAAAVA 2997

Qy        209 DLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLD---- 264
Db       2998 DLVA---RHRISAV---VHT--AGVLDDGVVESL-----TPERL---SAVLRPKVDAAWN 3041

Qy        265 ----TR--ELDAAYWYRNMSSTVRFE----------------------PAARLLL--- 291
Db       3042 LHEATRDLDLDAFVVFSSVAGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPW 3101

Qy        292 -----QQGPKTFVEMS--------PHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQG 338
Db       3102 SRDGGMTGTLTDVDSSASPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASL-RAQG 3160

Qy        339 TLDHFLTSLAQLRGH----GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDG 394
Db       3161 EVPPLLRGLIRTRARRSAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHAN 3220

Qy        395 NERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTA 454
Db       3221 PETIES----TRVFQDLGFDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDE 3276

Qy        455 ALGHLDEDTAPVP-DSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAI 513
Db       3277 LFGAQEEAELPAPVPSPAG-----AADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAV 3331

Qy        514 GPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRL 573
Db       3332 SDFPTDRGWDVESLYSPDPEALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRL 3391

Qy        574 LLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTG 633
Db       3392 LLETTWEAIERTGIDPASLRGSRTGVFAGVMYTDYGDLL------VGDQFEGYRSNGSAA 3445

Qy        634 SVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMF 693
Db       3446 SIASGRVSYTFGFEGPAVTVDTACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTF 3505

Qy        694 IEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQ 753
Db       3506 VEFSRQRGLSADGRCKAFADAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQ 3565

Qy        754 DGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ 813
Db       3566 DGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQ 3625

Qy        814 -RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGA 872
Db       3626 ERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGA 3685

Qy        873 VRLLTEAVPWP-GDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGL 931
Db       3686 VELLTSEAEWPQGEGP---RRAGVSSFGVSGTNAHVILEQ----------------PG- 3724

Qy        932 AISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVS 991
Db       3725 ----------PDAADA--APDATVTDPGALAWVLSARNEAALRCQAARLLSLVAGSDALC 3772

Qy        992 LADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVF 1051
Db       3773 ARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEADAGLAEG-ASGA--GRTAF 3829

Qy       1052 VFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVW-ERAD 1110
Db       3830 LFSGQGSQRLGMGWELYARYPVFADAFDAVCAALD--------EHLERPLRDVVWGEDAE 3881

Qy       1111 VV------QPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVA 1164
Db       3882 LLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGVFSLPDACALVA 3941

Qy       1165 LRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEV 1224
Db       3942 ARGRLMQQLPSGGAMMAIRATEDEVLPHLAE----GVSLAAVNGPSSVVISGAEDAVLAI 3997

Qy       1225 LAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTT 1284
Db       3998 AAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVVSNLTGRLAEPE 4057

Qy       1285 TL-DAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAI 1343
Db       4058 QLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMARESASD---DAVLA 4114

Qy       1344 GSLRRGDNDTRRFLTALAHTHTTG---------IGTPTTWHHHYTHHHTHPHPHTHLDLP 1394
Db       4115 PVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARW----------------ADLP 4158

Qy       1395 TYPFQHQHYWLESSQPGAGSGSGAGAGS------GAGSGRAGTAG----GTAEVESRFW- 1443
Db       4159 TYAFQHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWL 4218

Qy       1444 ------DAVARQDLETVATTLAVPPSAGLD----------TVVPALSAWHRHQ------- 1480
Db       4219 ADHVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD 4278

Qy       1481 ---------HDQARINTWTYQET----------------WKP------------------ 1497
Db       4279 EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLADA 4338

Qy       1498 -LTL-PTTHQPHQTWLIA--------IPETQTHHPHITNILTNLHHHGITPIPLTLNHTH 1547
Db       4339 GLTYGPAFHGLQAAWKLGGDVYAEAKLPES----------TDGDAYGL----------- 4376

Qy       1548 TNPQHLHHTLHHTRQQAQN----------HTTGA---------ITGLLSLLALDETP 1585
Db       4377 -HPALFDAALHASALGGAEAGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTS- 4434

Qy       1586 HPHHPHTPTGTLLNLTLTQTHTQTHP----------PTPLWYATTNATTTHPNDPL-THP 1634
Db       4435 --GAPVASVESLVIRPLSAGQVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYA 4492

Qy       1635 TQAQTWGLA--RTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQT-QLAIRTTG 1691
Db       4493 DLDSLEGAAVPGTVLVAPPSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRG 4552

Qy       1692 T--------------------------------HTRRLTPTTLTPTHQPP--------- 1709
```



```
Qy       3796 DNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVE 3855
Db       6615 DTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEIDQDRAFLDLGFDSLTAVE 6674

Qy       3856 LRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSAL 3915
Db       6675 LRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPTVGPEALLGELERMEKSFGGL 6734

Qy       3916 DKTDSASERVTLRLKSLMLRWNAPQHPTAES------ADDDEKFTSATEAEIFKFIDNDL 3969
Db       6735 DLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPSDEDFDFESASDDEVFDLLDNEL 6794

Qy       3970 GLS 3972
Db       6795 GLS 6797


RESULT 13
Q93HJ4
ID   Q93HJ4      PRELIMINARY;      PRT;  4840 AA.
AC   Q93HJ4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   OlmA2 protein.
GN   OLMA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69193.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   4840 AA;  504325 MW;  169BDA3589AD284F CRC64;

  Query Match             34.6%;  Score 7217;  DB 2;  Length 4840;
  Best Local Similarity   39.3%;  Pred. No. 1.5e-301;
  Matches 1798;  Conservative  475;  Mismatches 1419;  Indels  878;  Gaps   97;

Qy          6 GGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFA 65
Db        542 GEPDPRVVRGRTV-----PSGPVAPLFTGQGSQRAGIGRELYETFPVFARAFDEVRDRLD 596

Qy         66 PYVDWSVEQVLRDSPDAPGLDRVDVVQPT------LFAVMISLAALWRSQGVEPCAVLGH 119
Db        597 PQLAHPLDDIL----DGRVVDGLPLIEQTGYALAAVFALEVALFRLVEHAGVVPDFVLGH 652

Qy        120 SLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN 179
```

```
Db        653 STGELAAAHCAGVLSLDDACTLVAARGRFVQTLRTDGAMVAVETSETELAELIA--RADV 710

Qy        180 PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHS----PLMYAIE 235
Db        711 GDRATIAVVNGPRSVVVSGDADAVERVRARCEERGRKTRKLPISHAAHSVHMEPMLARLR 770

Qy        236 ERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQG- 294
Db        771 E-VAEGL---EHRPPRIPMVSNLTGAPVPPEANWPEYWVRQVRETVRFHDGIRWLGAQGV 826

Qy        295 ----------------------------------------PKTFV----EMSPHPVL---- 307
Db        827 TACLELGPGGALIAMARESLGDPGREVALVPTLQHDRPEPATFLSALAELHVHGVAVDWD 886

Qy        308 ----------TMGLQE------------------------------------------- 313
Db        887 TPLAARGATSTDLPTYAFQRERYWLPATPAAPRTAEPADRTGRVRYRVGWTPYPDVDTDA 946

Qy        314 --------LAPDLGDTTG--------------------TA------------DTVIMGT 332
Db        947 RPTGTWLVVAPD-GDSDGDGHGDGARVAEALGERAVLLTAGPGLAPALARLPEDRPVAGV 1005

Qy        333 LR------------------RGQGT----------------------------------- 339
Db       1006 LALPAESATRWTRLAPRSALRGRGARPASVRHQAPVSVDTEAVPGAAPALLWGLGRVPHS 1065

Qy        340 ------------------------------------------------------------ 339
Db       1066 GGGVVDLPAGLDARSLRLLAAVLAAPGDQDELAVRPGGVFARKLLPAQRPRRQRAWRTGG 1125

Qy        340 ------------------------------------------------------------ 339
Db       1126 TVLITGDIAGVTPELLQYLSGGGARTLLLAHRPGTPLPDAAVLAAAGGPCTVTEWDLGPL 1185

Qy        340 ----------------------------------------------------LDHF---- 343
Db       1186 ETASVTAVVHLSGIDPDGPSTPDALTAAGPARAVQDRLRTVVRLEELFHDADLDAFVLLS 1245

Qy        344 --------------------LTSLAQLR------------GHGETSATT----------- 360
Db       1246 TVAGTWGGAEDLVHTVAHAALESLAERRTLDGRPGTCVGWGPWETPADTGASPVPGLLPL 1305

Qy        361 ----VLSARLTALS------------------------------------------PTQQQ 375
Db       1306 PADLAVSALAEALDADDPVCAVVDVDWSLFHPVLTGRGPRPLVADLPEVRALPAPPTAEV 1365

Qy        376 SL---------LLDLVRAHTMAVLNDDGNERTASDAG--PSASFAHLGFDSVMGVELRNR 424
Db       1366 VLAPAADREHQLLELVRAQAAAVLGH------ASGAAIDPVRPFRDLGFESLAAVRFRDR 1419

Qy        425 LSKATGLRLPVTLIFDHTTPAAVAARLRTAALG---HLDEDTAPVPDSPSGHGGTAAADD 481
Db       1420 VAAETGLRLPATLVFDHPTAQAVAAHLLAELSGVRQHEDE-VAP----------TADHDD 1468

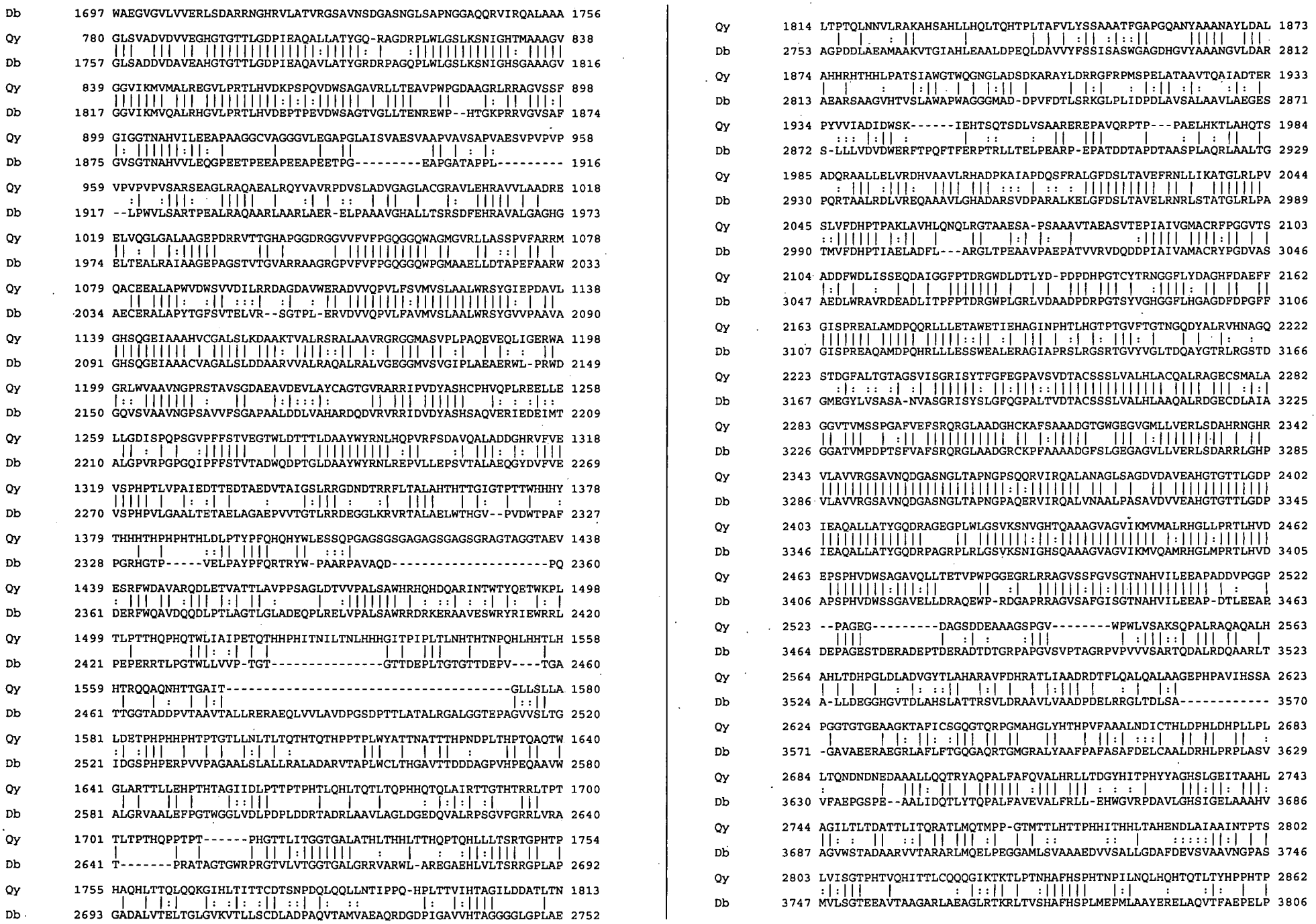
Qy        482 PIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGW--PTEQRHAQDPTQPGTFY 539
Db       1469 PIAIVGMACRYPGGVRGPEDLWDLVHCGRDGVGDFPTDRGWDLPALRRLAPD-------L 1521

Qy        540 PQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGV 599
Db       1522 AMKGGFLPDAADFDAAFFGISPREATAMDPQQRLLLEVSWQALEGAGLDPAALRGSRTGV 1581

Qy        600 FAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSA 659
Db       1582 FAGAAGSDYGVVL-----EGAPEADGYLTTGTAASVISGRIAYTLGLHGPAVTVDTACSS 1636

Qy        660 SLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTG 719
Db       1637 SLVALHMAVGALHQGECDLALATGVSVISTPGAFVDFARQSGLAADGRCKAFAASADGTN 1696

Qy        720 WGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANA 779
```



```
Qy       2863 LITANT----PPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTH 2918
Db       3807 VVSNVTGKPATAEQLCDPAYWARQVRGTVRFADGVDSLLGEEVTTFLELGPDGVLTAMAK 3866

Qy       2919 HNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQP-HTH----THLDL 2973
Db       3867 ESAGAGATGIAVQRRDRDEVRT-LLTALGRL------HVRTGRVDWPAFFHGTGARRVEL 3919

Qy       2974 PTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPW 3033
Db       3920 PTYAFQHRRYWLDAPTGGVEALAGAGLAGTGHPLLTAAVPLPGTGGSLFSGSL-----PG 3974

Qy       3034 LADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD 3093
Db       3975 IDEERPP-----SRGDILDLVLWAAGSLGCDRIAELDV--PAALPGGPHAPLQLVVAAPD 4027

Qy       3094 GEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAV 3153
Db       4028 EDGDRDFTLHV---GPTDSGDAGDESWTRIAHGTVGGAAAR---PEPDPAAWPPADAEPV 4081

Qy       3154 DVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGV 3213
Db       4082 GAD--------------------AVWRRADELFAELVLTEQEAGDADRFGLHPELLAGL 4120

Qy       3214 VQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVS 3273
Db       4121 TDLITELA------------GE----PVHFAGVTRYATGATELRVRLTRSGPG----AVA 4160

Qy       3274 VVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRG--SLYAVQWAEVGPV-PVCGQA 3330
Db       4161 VLLSDAEGEPVLSVDRVGVR---TGTAPATTPAGAPTGPDSLYRLTWTPVAAVAPEPAGP 4217

Qy       3331 WA--------------------WHEDVGESG----GGPVPGVVVLRCPDAGAGGGGGGGG 3366
Db       4218 WAVVGADGLAKALVATQGAEVEAHPDLASLGRSTEHGAAPGLVVLSVDTE------PGAT 4271

Qy       3367 GGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDG-PVDVVGASVWGLVRS 3425
Db       4272 VESMRHTTHRTLELVKELLADGRLAGARFVLVTRAAVSTG--DGTPVDPAQAAARGLLLS 4329

Qy       3426 AQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARL 3485
Db       4330 AQAEHPERFVLADLG---GRDADAELLA------AAVGTALTEAEPQLAVRDGQVLVPRL 4380

Qy       3486 KRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGT 3545
Db       4381 AQVTE--------------RPTDGPAERGPG----------GEH-------GTVVVTGAT 4409

Qy       3546 GVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREV 3605
Db       4410 GGIGTEIVRHLATAHGVRHLLLLSRRGPDDPRAAELGRVLAEAGADATFTACDTADRAAL 4469

Qy       3606 VRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELS 3665
Db       4470 ADALADVPAEHPVTAVVHIAGVVDDGVLTTLTPDRIDTVLRPKAEAAQHLHELTADLGLS 4529

Qy       3666 AFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLA 3725
Db       4530 HFVLFSSGVGTFGGAGQANYAAANAFLDALAHRRRAADLPATSLAWGLWETTAGMSAHLS 4589

Qy       3726 GTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPA---- 3781
Db       4590 DVDRRRMAQAGVLAFTPRQGLALFDAAWNSDATTLVPMRLDKAVLRGKAAEATLPAPFRG 4649

Qy       3782 -TRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840
Db       4650 LVRAPLRRASAAGARAAGQSFAQQLAEQSESGRRQLLLDLIHRQVAAVLDYAADAPLDAK 4709

Qy       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900
Db       4710 LPFRELGFDSLTAVELRNGLIAATGVQLSAALVFDYPTATALAEHLESKVLRSVAGAPLP 4769

Qy       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960
Db       4770 ALTQLDHLEAALATSPPDRATREQLAARLRRLASAWSADE-LEADDGDITSKLDSASDEE 4828

Qy       3961 IFKFIDNDLG 3970
Db       4829 LFDFINAEFG 4838
```

RESULT 14
Q9S0R3
ID Q9S0R3 PRELIMINARY; PRT; 4881 AA.
AC Q9S0R3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 4.
GN AVEA4.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84479.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 4881 AA; 510303 MW; F6568C1F01ACD37A CRC64;

Query Match 34.5%; Score 7190.5; DB 2; Length 4881;
Best Local Similarity 37.2%; Pred. No. 2.1e-300;
Matches 1750; Conservative 386; Mismatches 1069; Indels 1501; Gaps 71;

```
Qy          6 GGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESV-RACEAA- 63
Db        558 GG---RPVPGVVEG--RTTSGELAVLFAGQGTQRAGMGRELYEAYPVFAQAIDEICAEAD 612

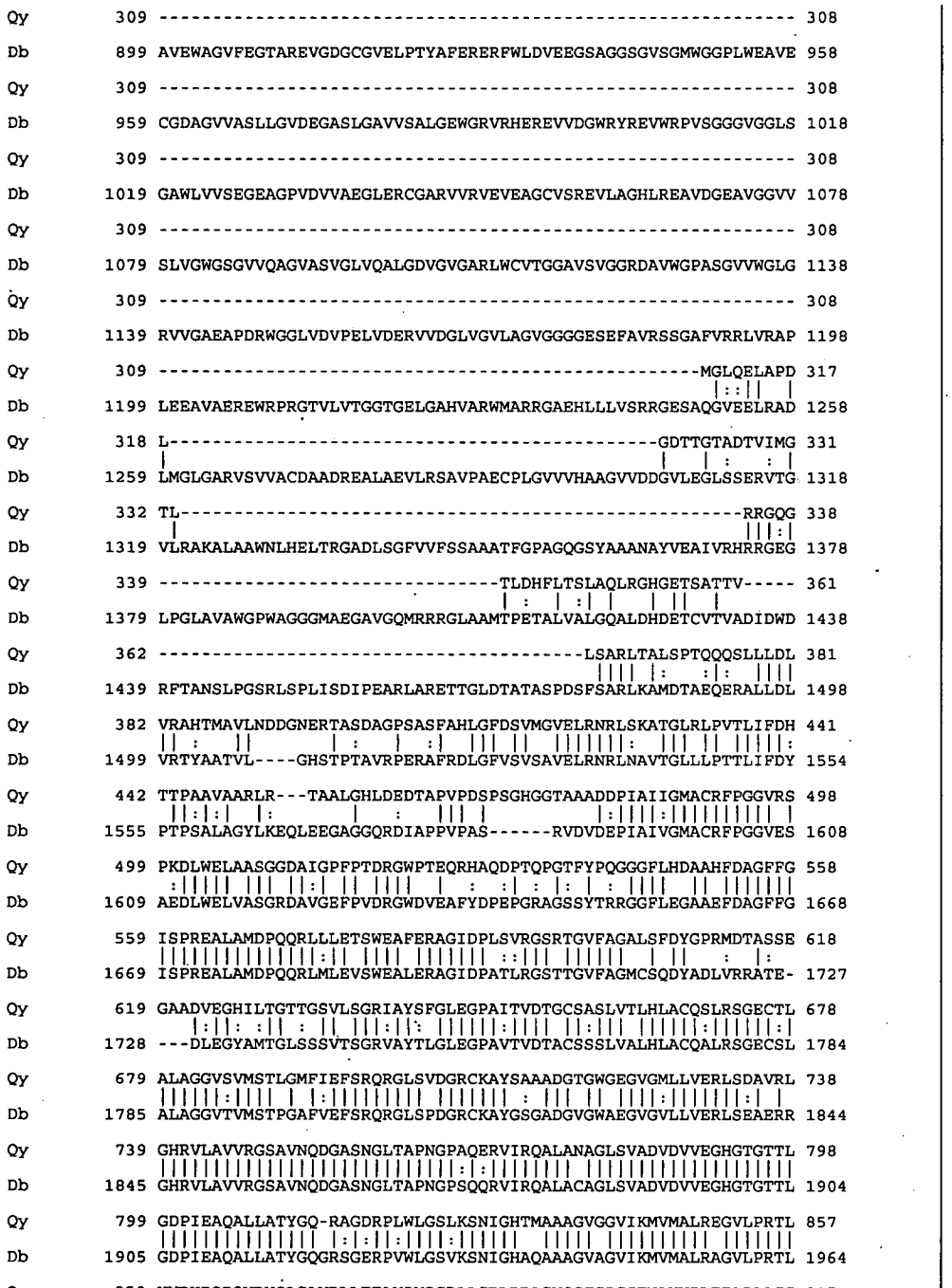
Qy         64 --------------FAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRS 108
Db        613 TARTDPGAPGLRDVLFAP----------QDSPEGRLIEDTGFAQPALFAFEVALFRLLET 662

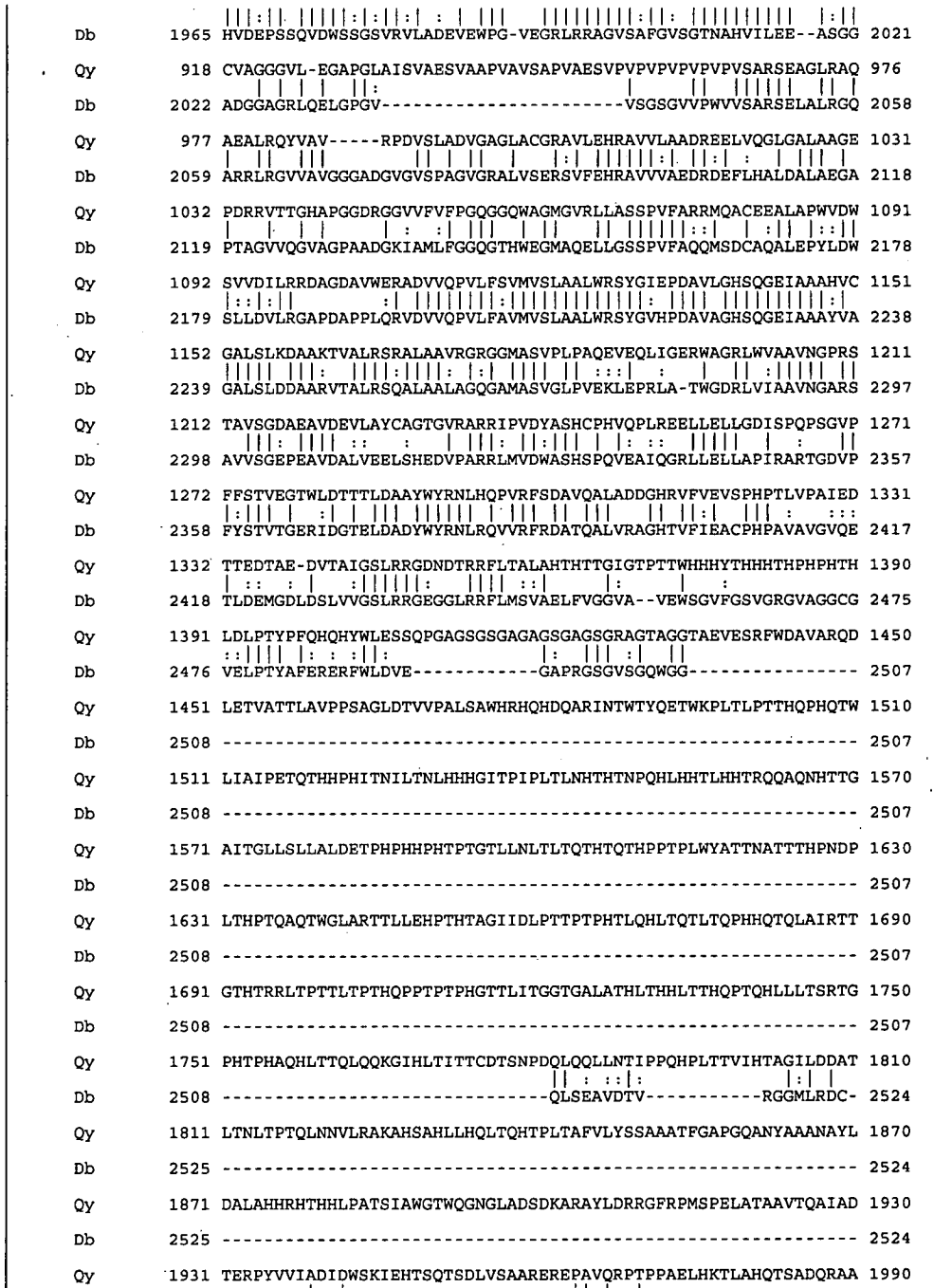
Qy        109 QGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDEL 168
Db        663 WGLTPDYVLGHSVGELAAAHVAGMLCLADAVALVVARGRLMQGLPSGGAMVAIEASEDEI 722

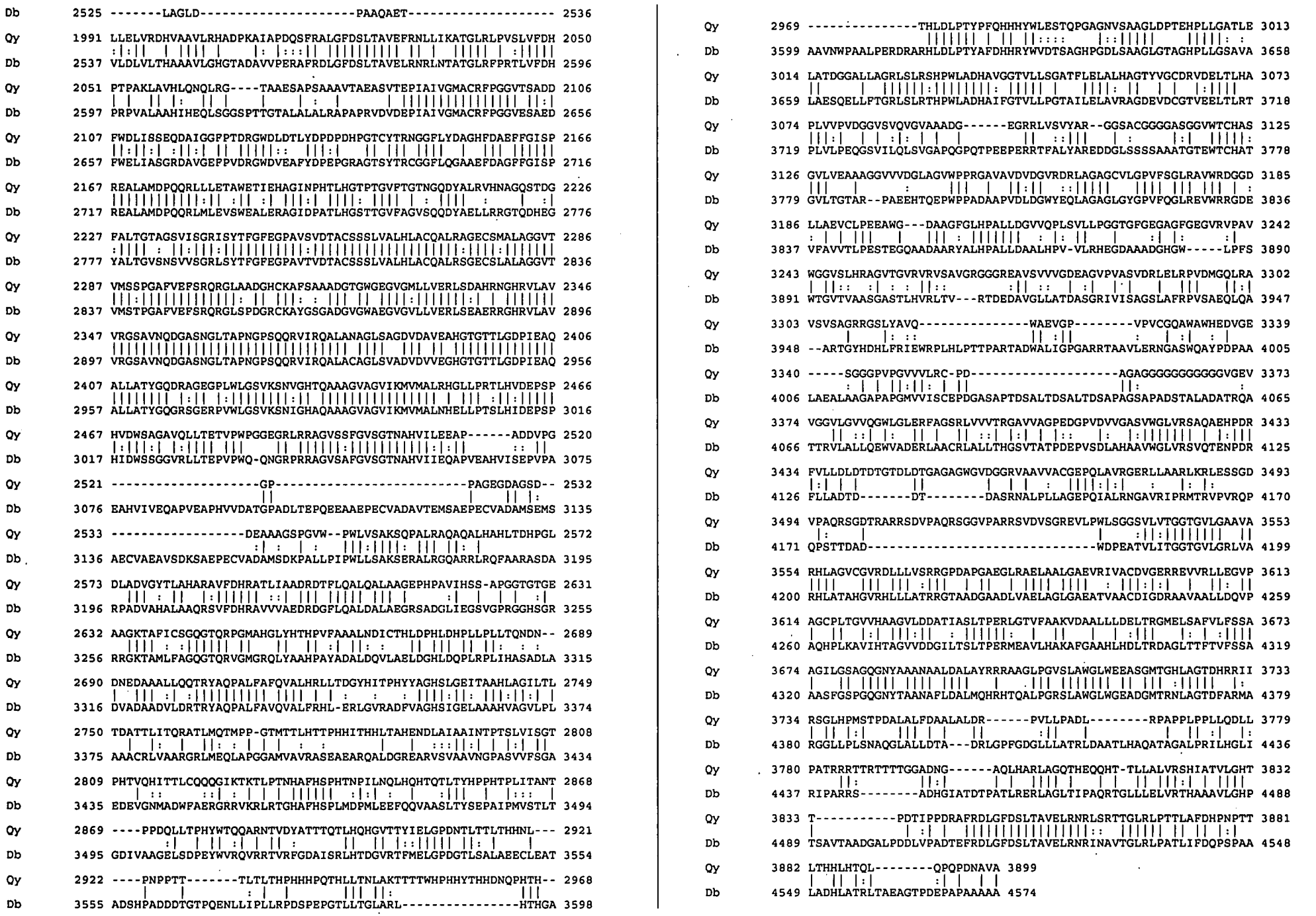
Qy        169 LPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHS 228
Db        723 L----PLPDEYASRVAHAAVNGPRSIVLSGDEDAVLDLAQQWAARGRRTRRLRTSHAFHS 778

Qy        229 PLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTREL-DAAYWYRNMSSTVRFEPAA 287
Db        779 PHMDAMLGDFRRAAEQVTFSAPRIPVVSNVTGAPLPAETMCTPDYWVEHARSTVRFADGI 838

Qy        288 RLLLQQGPKTFVEMSPHPVLT---------------------------------------- 308
Db        839 SWLQEQGVTTCLEIGPDGTLSALAQDSLSAPARAIPALRPDQPEARSVMTALAELFVAGT 898
```


Qy 309 ---------- 308
Db 899 AVEWAGVFEGTAREVGDGCGVELPTYAFERERFWLDVEEGSAGGSGVSGMWGGPLWEAVE 958
Qy 309 ---------- 308
Db 959 CGDAGVVASLLGVDEGASLGAVVSALGEWGRVRHEREVVDGWRYREVWRPVSGGGVGGLS 1018
Qy 309 ---------- 308
Db 1019 GAWLVVSEGEAGPVDVVAEGLERCGARVVRVEVEAGCVSREVLAGHLREAVDGEAVGGVV 1078
Qy 309 ---------- 308
Db 1079 SLVGWGSGVVQAGVASVGLVQALGDVGVGARLWCVTGGAVSVGGRDAVWGPASGVVWGLG 1138
Qy 309 ---------- 308
Db 1139 RVVGAEAPDRWGGLVDVPELVDERVVDGLVGVLAGVGGGGESEFAVRSSGAFVRRLVRAP 1198
Qy 309 ----------MGLQELAPD 317
Db 1199 LEEAVAEREWRPRGTVLVTGGTGELGAHVARWMARRGAEHLLLVSRRGESAQGVEELRAD 1258
Qy 318 L----------GDTTGTADTVIMG 331
Db 1259 LMGLGARVSVVACDAADREALAEVLRSAVPAECPLGVVVHAAGVVDDGVLEGLSSERVTG 1318
Qy 332 TL----------RRGQG 338
Db 1319 VLRAKALAAWNLHELTRGADLSGFVVFSSAAATFGPAGQGSYAAANAYVEAIVRHRRGEG 1378
Qy 339 ----------TLDHFLTSLAQLRGHGETSATTV----- 361
Db 1379 LPGLAVAWGPWAGGGMAEGAVGQMRRRGLAAMTPETALVALGQALDHDETCVTVADIDWD 1438
Qy 362 ----------LSARLTALSPTQQQSLLLDL 381
Db 1439 RFTANSLPGSRLSPLISDIPEARLARETTGLDTATASPDSFSARLKAMDTAEQERALLDL 1498
Qy 382 VRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDH 441
Db 1499 VRTYAATVL----GHSTPTAVRPERAFRDLGFVSVSAVELRNRLNAVTGLLLPTTLIFDY 1554
Qy 442 TTPAAVAARLR---TAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRS 498
Db 1555 PTPSALAGYLKEQLEEGAGGQRDIAPPVPAS------RVDVDEPIAIVGMACRFPGGVES 1608
Qy 499 PKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFG 558
Db 1609 AEDLWELVASGRDAVGEFPVDRGWDVEAFYDPEPGRAGSSYTRRGGFLEGAAEFDAGFFG 1668
Qy 559 ISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSE 618
Db 1669 ISPREALAMDPQQRLMLEVSWEALERAGIDPATLRGSTTGVFAGMCSQDYADLVRRATE- 1727
Qy 619 GAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTL 678
Db 1728 ---DLEGYAMTGLSSSVTSGRVAYTLGLEGPAVTVDTACSSSLVALHLACQALRSGECSL 1784
Qy 679 ALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRL 738
Db 1785 ALAGGVTVMSTPGAFVEFSRQRGLSPDGRCKAYGSGADGVGWAEGVGVLLVERLSEAERR 1844
Qy 739 GHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL 798
Db 1845 GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALACAGLSVADVDVVEGHGTGTTL 1904
Qy 799 GDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTL 857
Db 1905 GDPIEAQALLATYGQGRSGERPVWLGSVKSNIGHAQAAAGVAGVIKMVMALRAGVLPRTL 1964
Qy 858 HVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGG 917


Db 1965 HVDEPSSQVDWSSGSVRVLADEVEWPG-VEGRLRRAGVSAFGVSGTNAHVILEE--ASGG 2021
Qy 918 CVAGGGVL-EGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQ 976
Db 2022 ADGGAGRLQELGPGV----------VSGSGVVPWVVSARSELALRGQ 2058
Qy 977 AEALRQYVAV-----RPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGE 1031
Db 2059 ARRLRGVVAVGGGADGVGVSPAGVGRALVSERSVFEHRAVVVAEDRDEFLHALDALAEGA 2118
Qy 1032 PDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDW 1091
Db 2119 PTAGVVQGVAGPAADGKIAMLFGGQGTHWEGMAQELLGSSPVFAQQMSDCAQALEPYLDW 2178
Qy 1092 SVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVC 1151
Db 2179 SLLDVLRGAPDAPPLQRVDVVQPVLFAVMVSLAALWRSYGVHPDAVAGHSQGEIAAAYVA 2238
Qy 1152 GALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRS 1211
Db 2239 GALSLDDAARVTALRSQALAALAGQGAMASVGLPVEKLEPRLA-TWGDRLVIAAVNGARS 2297
Qy 1212 TAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVP 1271
Db 2298 AVVSGEPEAVDALVEELSHEDVPARRLMVDWASHSPQVEAIQGRLLELLAPIRARTGDVP 2357
Qy 1272 FFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIED 1331
Db 2358 FYSTVTGERIDGTELDADYWYRNLRQVVRFRDATQALVRAGHTVFIEACPHPAVAVGVQE 2417
Qy 1332 TTEDTAE-DVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTH 1390
Db 2418 TLDEMGDLDSLVVGSLRRGEGGLRRFLMSVAELFVGGVA--VEWSGVFGSVGRGVAGGCG 2475
Qy 1391 LDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQD 1450
Db 2476 VELPTYAFERERFWLDVE----------GAPRGSGVSGQWGG---------- 2507
Qy 1451 LETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTW 1510
Db 2508 ---------- 2507
Qy 1511 LIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG 1570
Db 2508 ---------- 2507
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2508 ---------- 2507
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2508 ---------- 2507
Qy 1691 GTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTG 1750
Db 2508 ---------- 2507
Qy 1751 PHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDAT 1810
Db 2508 ----------QLSEAVDTV----------RGGMLRDC- 2524
Qy 1811 LTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYL 1870
Db 2525 ---------- 2524
Qy 1871 DALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIAD 1930
Db 2525 ---------- 2524
Qy 1931 TERPYVVIADIDWSKIEHTSQTSDLVSAARREPAVQRPTPPAELHKTLAHQTSADQRAA 1990



```
RESULT 15
Q93NW6
ID   Q93NW6      PRELIMINARY;      PRT; 10917 AA.
AC   Q93NW6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphC.
GN   AMPHC.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73514.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00059; ADH_ZINC; UNKNOWN_1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
KW   Phosphopantetheine.
SQ   SEQUENCE   10917 AA;  1132905 MW;  15AC5956BB5810A1 CRC64;

  Query Match             34.4%;  Score 7168;  DB 2;  Length 10917;
  Best Local Similarity   41.5%;  Pred. No. 5.9e-299;
  Matches 1742;  Conservative  475;  Mismatches 1374;  Indels  603;  Gaps   93;

Qy        154 GTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVV----SGAREAVADLVAD 209
Db       6952 GTGGLGAVLAR------------------HLVAEHGVRELVLVSRRGGAAAGAAELVAE 6992

Qy        210 LTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGL----------LPITPRPSRIPFHSSVT 259
Db       6993 LAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGLIASLTPERL---SGVL 7049

Qy        260 GGRLD--------TRELD------------------AAYWYRN--MSSTVRFEPAA--- 287
Db       7050 RPKVDAAWNLHEATRDLDLDAFVVFSSVAGVFGNAGQANYAAGNAFLDALMEHRRAAGLP 7109

Qy        288 RLLLQQGP----------KTFVEMSPHPVLTMGLQELAPDLG----DTTGTADTVIMGTL 333
Db       7110 GLSLAWGPWEQTGTGGMTGTLSDAELERLASSGVPPLAVEQGLALFDAALSTEDAALVPV 7169

Qy        334 R------RGQGTLDHFLTSLAQLRGH-----GETSATTVLSARLTALSPTQQQSLLLDLV 382
Db       7170 RLDLPALRGQGEVPPLLRSLIRARSRRAAVAGSATATG-LRDRLAVLDAVERREVLVDLV 7228

Qy        383 RAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHT 442
Db       7229 RGQVALVL----GHADATAVHPTRAFRDLGFDSLTSVELRNRLNAVTGLRLPATMVFDYP 7284

Qy        443 TPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDL 502
Db       7285 TVEVLVDFILDELLG--TEVEAEVVQ----RGTAAVADDPIVIVGMACRYPGGVASPEDL 7338

Qy        503 WELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPR 562
Db       7339 WRLVTEGTDAVSAFPVNRGWDVENLYHPDPDNIGTSYTRSGGFLHEAGEFDPGFFGMSPR 7398

Qy        563 EALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAAD 622
Db       7399 EALSTDSQQRLLLETSWEAIERAGIDPVGLRGSATGVFAGVMYSDYSAML------ASPE 7452

Qy        623 VEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG 682
Db       7453 FEGFQGSGSSPSVASGRVSYTFGFEGPAVTVDTACSSSLVAMHWAMQALRSGEISLALAG 7512

Qy        683 GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRV 742
Db       7513 GVTVMSTPGVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLERQSDAIRNGHEI 7572

Qy        743 LAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPI 802
Db       7573 LAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAQEIDVVEAHGTGTTLGDPI 7632

Qy        803 EAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDK 861
Db       7633 EAQALLATYGRDRDPEQPLWLGSVKSNIGHTQAAAGVAGVIKMVMSMRHGVLPRTLHVDA 7692

Qy        862 PSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEE-APAAGGCVA 920
Db       7693 PSSHVDWTEGAVELLTEQTEWP--ETEHVRRAGVSSFGISGTNAHVILEQPAKVIQGTVI 7750

Qy        921 GGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEAL 980
Db       7751 GGSTPESG------------------------VVEPSVVPWVLSGKTPEALRSQAAKL 7784

Qy        981 RQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGH 1040
Db       7785 LASVEAELDRPLVDVGSSLVAARSLFEHRAVVLATDADTAARALAALAVGEPDPAAVSGP 7844

Qy       1041 APGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDIL-RR 1099
Db       7845 ARTGRSAA---LFSGQGSQRLGMGRELYGRFPVFAEALDAVLAVLDGELDGSLREVMWGE 7901

Qy       1100 DAG---DAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSL 1156
Db       7902 DAGLLNETGW-----TQPALFAVEVALHRLVESFGVTPDFVAGHSIGEIAAAHIAGVFSL 7956

Qy       1157 KDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSG 1216
Db       7957 EDAARLVAARGRLMQALPAGGAMVAVQATEDEVIPYLSD----EVSIAALNGPASVVVSG 8012

Qy       1217 DAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTV 1276
Db       8013 AEGAVLEVAARFEAEGRKATMLRVSHAFHSPLMDPMLDEFRAVAEGLSFAAPVIPVVSNL 8072

Qy       1277 EGTWLDTTTL-DAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTED 1335
Db       8073 TGSLATAEELCSPEYWVRHVREAVRFADGVTTLEEQGVTTFLELGPDGVLSAMAQESL-- 8130

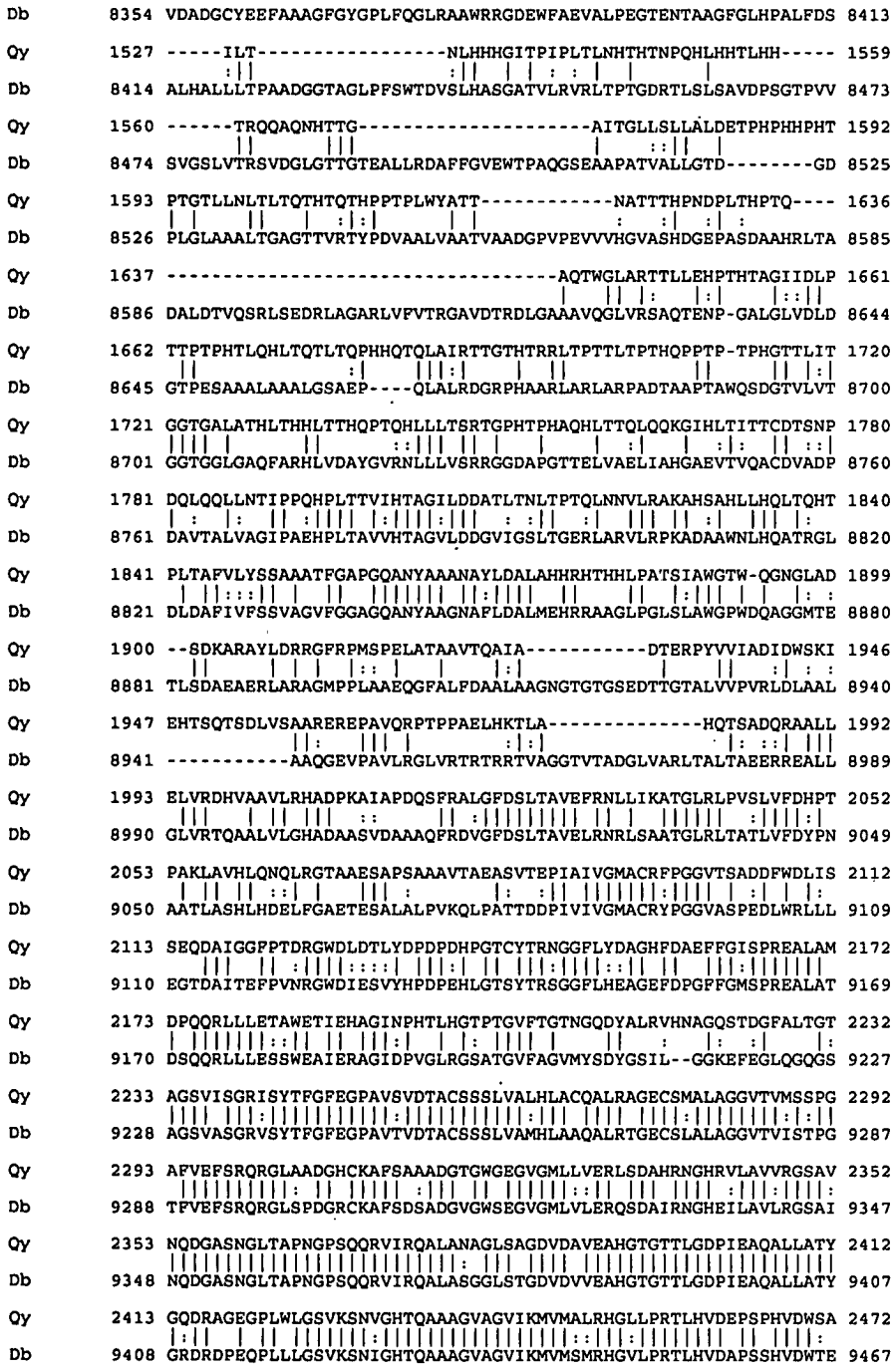
Qy       1336 TAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGI---------GTPTTWHHHYTHHHTHPH 1386
Db       8131 TGDEAVTVPLLRKNTAEEAATVAALARLHIGGLRIDWSALFAGTGA-------------- 8176

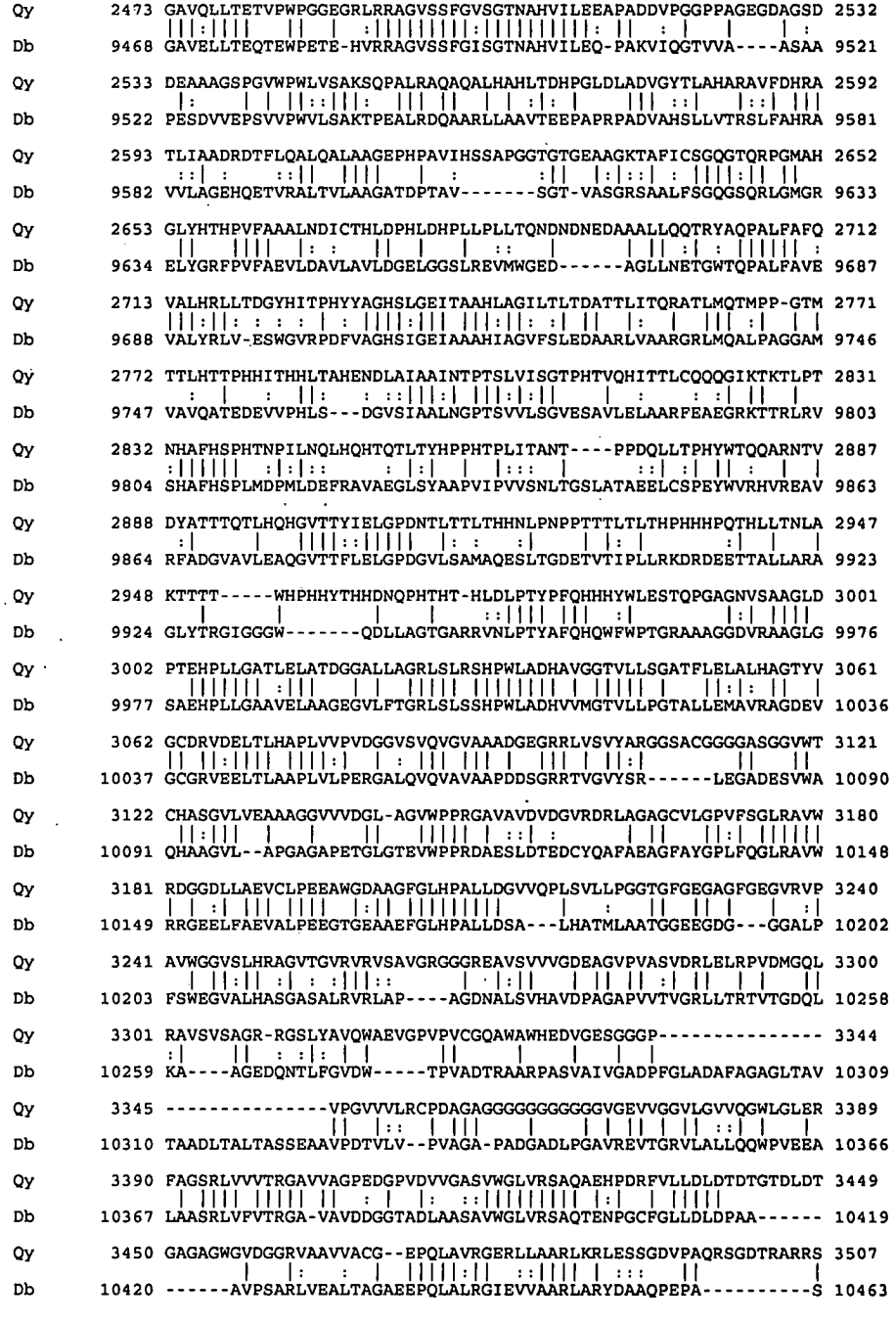
Qy       1387 PHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGS-----------GAGSGRAGTA-- 1432
Db       8177 --RRVELPTYAFQHQWFW-PAAPLGGGDVRAAGLGSAEHPLLGAAVELAAGEGVLFTGRL 8233

Qy       1433 ------------------GGTA--EVESRFWDAVARQDLE--TVATTLAVP-------- 1461
Db       8234 ALHSHPWLADHTVDGTVLLPGTALLELAIRAGDEVGSDQVEELTITAPLVLPERGAVQVQ 8293

Qy       1462 -----PSAGLDTVVPAL--------SAWHRH----------------------QHDQA 1484
Db       8294 VAVEAPDAAGRRVVGVYARPEGTGDTAWSQHAAGVLAPAGSRPAPAAFDTAVWPPEGAQA 8353

Qy       1485 RINTWTYQE------TWKPL--TLPTTHQPHQTWL--IAIPETQTH-------HPHITN- 1526
```


Db 8354 VDADGCYEEFAAAGFGYGPLFQGLRAAWRRGDEWFAEVALPEGTENTAAGFGLHPALFDS 8413
Qy 1527 -----ILT----------NLHHHGITPIPLTLNHTHTNPQHLHHTLHH----- 1559
Db 8414 ALHALLLTPAADGGTAGLPFSWTDVSLHASGATVLRVRLTPTGDRTLSLSAVDPSGTPVV 8473
Qy 1560 ------TRQQAQNHTTG----------AITGLLSLLALDETPHPHHPHT 1592
Db 8474 SVGSLVTRSVDGLGTTGTEALLRDAFFGVEWTPAQGSEAAPATVALLGTD--------GD 8525
Qy 1593 PTGTLLNLTLTQTHTQTHPPTPLWYATT----------NATTTHPNDPLTHPTQ---- 1636
Db 8526 PLGLAAALTGAGTTVRTYPDVAALVAATVAADGPVPEVVVHGVASHDGEPASDAAHRLTA 8585
Qy 1637 ----------AQTWGLARTTLLEHPTHTAGIIDLP 1661
Db 8586 DALDTVQSRLSEDRLAGARLVFVTRGAVDTRDLGAAAVQGLVRSAQTENP-GALGLVDLD 8644
Qy 1662 TTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTP-TPHGTTLIT 1720
Db 8645 GTPESAAALAAALGSAEP----QLALRDGRPHAARLARLARPADTAAPTAWQSDGTVLVT 8700
Qy 1721 GGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNP 1780
Db 8701 GGTGGLGAQFARHLVDAYGVRNLLLVSRRGGDAPGTTELVAELIAHGAEVTVQACDVADP 8760
Qy 1781 DQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHT 1840
Db 8761 DAVTALVAGIPAEHPLTAVVHTAGVLDDGVIGSLTGERLARVLRPKADAAWNLHQATRGL 8820
Qy 1841 PLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNGLAD 1899
Db 8821 DLDAFIVFSSVAGVFGGAGQANYAAGNAFLDALMEHRRAAGLPGLSLAWGPWDQAGGMTE 8880
Qy 1900 --SDKARAYLDRRGFRPMSPELATAAVTQAIA----------DTERPYVVIADIDWSKI 1946
Db 8881 TLSDAEAERLARAGMPPLAAEQGFALFDAALAAGNGTGTGSEDTTGTALVVPVRLDLAAL 8940
Qy 1947 EHTSQTSDLVSAARREREPAVQRPTPPAELHKTLA----------HQTSADQRAALL 1992
Db 8941 ----------AAQGEVPAVLRGLVRTRTRRTVAGGTVTADGLVARLTALTAEERREALL 8989
Qy 1993 ELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPT 2052
Db 8990 GLVRTQAALVLGHADAASVDAAAQFRDVGFDSLTAVELRNRLSAATGLRLTATLVFDYPN 9049
Qy 2053 PAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLIS 2112
Db 9050 AATLASHLHDELFGAETESALALPVKQLPATTDDPIVIVGMACRYPGGVASPEDLWRLLL 9109
Qy 2113 SEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAM 2172
Db 9110 EGTDAITEFPVNRGWDIESVYHPDPEHLGTSYTRSGGFLHEAGEFDPGFFGMSPREALAT 9169
Qy 2173 DPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGT 2232
Db 9170 DSQQRLLLESSWEAIERAGIDPVGLRGSATGVFAGVMYSDYGSIL--GGKEFEGLQGQGS 9227
Qy 2233 AGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPG 2292
Db 9228 AGSVASGRVSYTFGFEGPAVTVDTACSSSLVAMHLAAQALRTGECSLALAGGVTVISTPG 9287
Qy 2293 AFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAV 2352
Db 9288 TFVEFSRQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLERQSDAIRNGHEILAVLRGSAI 9347
Qy 2353 NQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATY 2412
Db 9348 NQDGASNGLTAPNGPSQQRVIRQALASGGLSTGDVDVVEAHGTGTTLGDPIEAQALLATY 9407
Qy 2413 GQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSA 2472
Db 9408 GRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVMSMRHGVLPRTLHVDAPSSHVDWTE 9467


Qy 2473 GAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSD 2532
Db 9468 GAVELLTEQTEWPETE-HVRRAGVSSFGISGTNAHVILEQ-PAKVIQGTVVA----ASAA 9521
Qy 2533 DEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRA 2592
Db 9522 PESDVVEPSVVPWVLSAKTPEALRDQAARLLAAVTEEPAPRPADVAHSLLVTRSLFAHRA 9581
Qy 2593 TLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH 2652
Db 9582 VVLAGEHQETVRALTVLAAGATDPTAV-------SGT-VASGRSAALFSGQGSQRLGMGR 9633
Qy 2653 GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ 2712
Db 9634 ELYGRFPVFAEVLDAVLAVLDGELGGSLREVMWGED------AGLLNETGWTQPALFAVE 9687
Qy 2713 VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPP-GTM 2771
Db 9688 VALYRLV-ESWGVRPDFVAGHSIGEIAAAHIAGVFSLEDAARLVAARGRLMQALPAGGAM 9746
Qy 2772 TTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPT 2831
Db 9747 VAVQATEDEVVPHLS---DGVSIAALNGPTSVVLSGVESAVLELAARFEAEGRKTTRLRV 9803
Qy 2832 NHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLTPHYWTQQARNTV 2887
Db 9804 SHAFHSPLMDPMLDEFRAVAEGLSYAAPVIPVVSNLTGSLATAEELCSPEYWVRHVREAV 9863
Qy 2888 DYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLA 2947
Db 9864 RFADGVAVLEAQGVTTFLELGPDGVLSAMAQESLTGDETVTIPLLRKDRDEETTALLARA 9923
Qy 2948 KTTTT-----WHPHHYTHHDNQPHTHT-HLDLPTYPFQHHHYWLESTQPGAGNVSAAGLD 3001
Db 9924 GLYTRGIGGGW-------QDLLAGTGARRVNLPTYAFQHQWFWPTGRAAAGGDVRAAGLG 9976
Qy 3002 PTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYV 3061
Db 9977 SAEHPLLGAAVELAAGEGVLFTGRLSLSSHPWLADHVVMGTVLLPGTALLEMAVRAGDEV 10036
Qy 3062 GCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWT 3121
Db 10037 GCGRVEELTLAAPLVLPERGALQVQVAVAAPDDSGRRTVGVYSR------LEGADESVWA 10090
Qy 3122 CHASGVLVEAAAGGVVVDGL-AGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180
Db 10091 QHAAGVL--APGAGAPETGLGTEVWPPRDAESLDTEDCYQAFAEAGFAYGPLFQGLRAVW 10148
Qy 3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
Db 10149 RRGEELFAEVALPEEGTGEAAEFGLHPALLDSA---LHATMLAATGGEEGDG---GGALP 10202
Qy 3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
Db 10203 FSWEGVALHASGASALRVRLAP----AGDNALSVHAVDPAGAPVVTVGRLLTRTVTGDQL 10258
Qy 3301 RAVSVSAGR-RGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGP--------------- 3344
Db 10259 KA----AGEDQNTLFGVDW-----TPVADTRAARPASVAIVGADPFGLADAFAGAGLTAV 10309
Qy 3345 ---------------VPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLER 3389
Db 10310 TAADLTALTASSEAAVPDTVLV--PVAGA-PADGADLPGAVREVTGRVLALLQQWPVEEA 10366
Qy 3390 FAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDT 3449
Db 10367 LAASRLVFVTRGA-VAVDDGGTADLAASAVWGLVRSAQTENPGCFGLLDLDPAA------ 10419
Qy 3450 GAGAGWGVDGGRVAAVVACG--EPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRS 3507
Db 10420 ------AVPSARLVEALTAGAEEPQLALRGIEVVAARLARYDAAQPEPA----------S 10463

```
Qy       3508 DVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLL 3567
Db      10464 GAMADGTSGT--------------FRTTGTVLVTGGTGGLGRLLARHLVTTHGVRSLLL 10508

Qy       3568 VSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGV 3627
Db      10509 ASRRGLQADGAGKLVAELEEMGAEVTVRACDVADPEAVAALVAAAPKGRPLSAVVHTAGV 10568

Qy       3628 LDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAA 3687
Db      10569 LDDGVVGSLTPERLDTVLRPKADAVWHLHRATRGLDLDAFVVYSSVAGVFGGAGQANYAA 10628

Qy       3688 ANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALA 3747
Db      10629 ANTFLDALMARRRADGLPGVSLAWGPWEQDSGMTGTLTDAEAERLARSGMPPLTVEQGHA 10688

Qy       3748 LFDAALALDRPVLLP-----ADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLH 3802
Db      10689 LFDAVVAGERALALPVRLDLAVLRAGGEVPALLRGLIRVPARRAAAQSAVPSADT---LT 10745

Qy       3803 ARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSR 3862
Db      10746 RQLAGLGAQEQEEALLTLVRGQAATVLGHRDGSAIGRERQFQELGFDSLTAVEFRNRLNT 10805

Qy       3863 TTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSAS 3922
Db      10806 ATGLRLPATLLFDYPTPADVVRYLRGQLVTDDPSGAGSVLAALEGLEKAIAALSLDDAEE 10865

Qy       3923 -ERVTLRLKSLMLRWNAPQHPTAESADDD---EKFTSATEAEIFKFIDNDLGLS 3972
Db      10866 HRRVAGRIEVLRTKWAALGR--TDTADEDASSDDLDGASDEDMFALLDDELGLS 10917
```

Search completed: June 17, 2003, 13:07:50
Job time : 304.577 secs